Embodied agents from a user's perspective

van Vugt, H.C.

2008

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
van Vugt, H. C. (2008). *Embodied agents from a user's perspective*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. Ridderprint.

Embodied agents from a user's perspective

Henriette van Vugt

The research reported in this dissertation has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems, and of NESCoR, the Dutch Research School for Communication Science.

The research is supported by a grant of the VU Research School for Business Information Sciences (VUBIS). VUBIS is a collaboration of the Faculty of Economics and Business Administration (FEWEB), the Faculty of Social Sciences (FSW) and the Faculty of Sciences (FEW) at the VU University Amsterdam, project title *Designing the user experience of Web as receivers of mass communication*.

Manuscriptcommissie:
prof.dr. Jan van Dijk (Universiteit van Twente)
dr. Jonathan Gratch (University of Southern California)
prof.dr. Nicole Krämer (Universität Duisburg-Essen)
prof.dr. Anton Nijholt (Universiteit van Twente)
prof.dr. Peter Vorderer (Vrije Universiteit Amsterdam)

ISBN 978-90-9023086-3
Printed by: Ridderprint, Offsetdrukkerij BV, Ridderkerk

SIKS Dissertation series No. 2008-16
Dissertation series faculty of Social Sciences, VU University Amsterdam No. 27

VRIJE UNIVERSITEIT

Embodied agents from a user's perspective

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Sociale Wetenschappen
op woensdag 25 juni 2008 om 10.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Henriette Christine van Vugt
geboren te Tilburg

promotoren:	prof.dr. J. Kleinnijenhuis
	prof.dr. G.C. van der Veer
copromotoren:	dr. J.F. Hoorn
	dr. E.A. Konijn

Contents

CHAPTER 1

Introduction

1.1 What elephants and interfaces have in common

'Nature does nothing without a purpose' (Aristotle, 384-322 B.C.)

At first sight, elephants and interfaces do not have much in common. Elephants are living creatures, whereas interfaces are part of computers (according to a design perspective, cf. Moran, 1981). Nevertheless, elephants and interfaces have certain things in common, as this introduction will explain. These shared attributes will serve as a starting-point for the study of user responses to *embodied agents*. Put simply, embodied agents are computer programs that have life-like visual representations and appear on a computer screen, such as on the Internet (e.g., http://www.ikea.com[1], in productivity applications (e.g., Clippy in Microsoft applications) or in computer games (e.g., Grand Theft Auto). The central question this dissertation examines is how computer users respond to embodied agents. Understanding user-agent interaction is useful for designers who want to use embodied agents in software products. In interface design, as in Aristotle's nature, nothing should be designed without a purpose.

In the animal kingdom, the *form* of an animal is crucial for its survival. For example, an elephant needs legs to move, and its legs are huge, thick pillars, necessary to support an elephant's bulk. Although an elephant cannot use pick up blades of grass or rip off branches off a tree with such hefty legs, its trunk, comprised of over forty thousand individual muscles, offers a solution. The form of an animal is related to its *function*. For example, in many animal species, the males hunt for food and protect their families against dangers in the environment, and therefore require muscles that are better developed than those of females, who manage the care

[1] Retrieved January 1st 2008

and feeding of the offspring. In elephant society, females spend their lives in family groups made up of mothers, daughters, sisters, and aunts, whereas males live more solitary lives[2]. Because only the most dominant males breed with cycling females, male elephants spend much more time fighting for dominance amongst themselves than the females do. Furthermore, an animal's *outer appearance* in terms of beauty is an indicator for an animal's strength and how well it can survive (cf. Etcoff, 1999). For example, male peacocks with beautiful tails have the best chances of mating and producing offspring: the more beautiful, the stronger, and the higher the likelihood of successful young. Thus form, function, and appearance aspects of animals influence their roles in the animal kingdom.

The same is true for humans. The form of our bodies enables us to move (legs), talk (tongue, vocal chords, etcetera), eat (teeth, stomach), and perform other bodily functions. More than for animals, however, humans assign certain functions to particular individuals. For example, a farmer grows food, a police officer maintains safety, a doctor provides medical help, a teacher transfers information and promotes the educational development of children. However, humans do not judge other people on rational grounds alone. Appearance also plays a role. The ancient Egyptians used eye liner to beautify themselves, and African tribes have a tradition of decorating their lips, ears, and necks. Nowadays, millions are spent in the fashion industry, and 'extreme make-over' programs are popular. Humans want to look good, and for good reasons. In general, an attractive person has a greater chance of getting an office job than an unattractive person, even if appearance has nothing to do with successfully performing that job (Watkins and Johnston, 2000; Cash et al., 1977; Dipboye et al., 1975). Even Aristotle observed that 'Beauty is a greater recommendation than any letter of introduction' (cited in Etcoff, 1999, p30). In short, form, function, and appearance of humans influence their interactions in society.

The importance of beauty did not emerge by accident and is no cultural construct, but is a result of evolution (e.g., Etcoff, 1999). What humans regard as beautiful or ugly, however, does depend on culture and time, which is why the fashion industry can influence what clothes we wear and what shoes we like. Even three- to six- month old babies spend more time looking at attractive than at non-attractive faces, and they play with attractive dolls for a longer time than with non-attractive dolls (e.g., Langlois et al., 1990, 1991). Thus, it is not surprising that humans judge inanimate objects not only in terms of their functional aspects (what can a person use the object for) but on appearance as well. For example, even though a car's function is merely to transport someone from one place to another, it will only sell well if it has a nice appearance. Likewise, a winter coat should be not only warm but fashionable as well, and a child will only enjoy playing with a doll if it looks like a clown, knight, or ballet dancer (i.e., it should be beautiful and have a realistic appearance as well).

In this dissertation, the research objects are *embodied agents* on computer screens. Functional aspects seem to be the primary concern for computer users.

[2]http://en.wikipedia.org/wiki/Elephant

What tasks can the embodied agent do, and can it perform those tasks *properly*? For example, is the agent able to search efficiently and effectively on the Internet? Does the agent offer useful advice? People will only ask an embodied agent for advice if the agent seems knowledgeable. Otherwise they will discard the agent into the virtual 'bin'. As long as the user is able to clearly communicate to the computer what tasks should be done, a dialogue window may suffice and the 'body' of an embodied agent seems of minor importance. The body, or form, of an embodied agent only seems important for task performance when it becomes functional. For example, a virtual salsa teacher needs legs to demonstrate steps. Or, if an agent wants to communicate with people who are hard of hearing, hands for sign language are indispensable. Most of the time, however, the form of an agent does not affect what actions it can perform on the computer. Whether or not an agent has legs does not affect its ability to search the Internet or give the user advice. Nevertheless, the agent's outer appearance may affect the user. A smiling face (or smiley ':-)') is appreciated more than a sad face (':-('), like in the 'real' world (e.g., Mueser et al., 1984). Or, a boy may prefer a game character mainly because of its heroic appearance and ignore a less heroic-looking character (Konijn et al., 2007). Central in this dissertation is the idea that user responses to embodied agents depend on both their function and outer appearance[3], just as animal or human responses to elephants, peacocks, humans, cars, winter coats, and dolls.

Yet, human responses are more complex than the above statements suggest. This can be illustrated by the following excerpt from Henry James' letter to his father about meeting the unmarried, 19th century novelist George Eliot (James and Edel, 1987, cited in Alam and Dover (2001)):

> 'She is magnificently ugly. She has a low forehead, dull gray eyes, a vast pendulous nose, a huge mouth, full of uneven teeth. Now in this vast ugliness resides a most powerful beauty, which, in a very few minutes, steals forth and charms the mind, so that you end as I ended, in falling in love with her.'

For Henry James, George Eliot's 'beauty' does not reside in her appearance, which most would consider to determine beauty. In addition, a beautiful appearance may not necessarily lead to happiness and fulfillment. A less beautiful peacock could have more offspring than a peacock with more beautiful plumage. The less physically attractive candidate of two may get the job. Someone may drive a car that he describes as plain ugly, for many years. A child's lovely new doll may disappear into the closet after only one day. Sometimes people prefer ugliness to beauty, and usefulness and beauty do not always produce positive attitudes and use. Therefore, we can assume that how users respond to embodied agents is also a complex matter.

[3]Form aspects can be considered part of appearance and will not be separately considered in this dissertation

Of course, the analogy between embodied agents, humans, and animals is not a complete one. In the end, embodied agents are not living creatures that develop naturally. Rather, they are virtual entities that humans design and create. Nevertheless, the above analogy illustrates that people's responses to embodied agents bears resemblance to the way humans respond to living creatures. What matters is not just whether the agent looks like a real human or animal (realism). Appearance aspects such as beauty, and more importantly, functional aspects (the agent's abilities) determine how people respond to the agent.

1.2 Embodied agent research

In computer science, the term 'agent' traditionally describes 'intelligent' software that is capable of flexible and autonomous action and behavior (Wooldridge and Jennings, 1995). Although not all computer programs with a human- or animal-like representation can be considered intelligent, it is common practice to use the terms 'embodied agent' or 'embodied conversational agent' to describe such programs (e.g., Cassell et al., 2000; Ruttkay and Pelachaud, 2004). Yet, researchers have used many other related terms as well, such as animated interface agents (e.g., Dehn and Van Mulken, 2000), intelligent (virtual) agents (for example, a yearly conference is called Intelligent Virtual Agents), lifelike characters (Prendinger and Ishizuka, 2004), animated characters (Rickenberg and Reeves, 2000), interface characters (Van Vugt et al., 2006), and virtual humans (cf. the Virtual Human Interaction Lab at Stanford University, and Swartout et al., 2006). The word *embodied* obviously refers to their visual, often human- or animal like, representation (cf. film characters or cartoons). Thus, 'embodied agent' is a generic term for different types of computer programs that have life-like visual appearances and appear on computer screens.

This section will focus on previous research examining user responses to embodied agents. First we provide several examples of embodied agents and clarify why systems may or may not improve by incorporating an embodied agent.

Embodied agents may play the role of sales assistant, customer service employee, teacher, storyteller, health advisor, or another role. For example, the virtual hostess Amber tries to answer customer questions on the ABNAMRO web site[4]. Or, the web site of the Dutch railway system[5], has a virtual coach available for download. The virtual coach, operated by a professional human trainer, provides the user with simple, general tips to stay healthy, such as, 'drink one glass of water every hour', or, 'get coffee for all your colleagues on your floor'. Another example in the health domain is the exercise advisor Laura, developed by Bickmore and Picard (2005). Using her social-emotional skills, Laura motivates people to do their daily physical exercises. Earlier, in 1997, Lester et al. (1997) developed embodied agents that tutored students. More recently, a research group developed

[4]http://www.abnamro.nl, Retrieved September 1st 2007
[5]http://www.nsensport.nl, Retrieved January 1st 2008

the ‘FearNot!’ system to teach children to deal with bullying behavior in schools (Aylett et al., 2007). In the entertainment domain, Mateas and Stern (2003) created the ‘Façade’ system in which users actively intervene in the marital problems of a successful and attractive couple in their early thirties, the virtual humans Grace and Trip. A well-known example from Microsoft is Bonzi Buddy, intended to make searching the Internet more fun, but often experienced as disruptive and annoying. Some even claim that Bonzi Buddy contains spy ware[6]. Thus, a large variety of embodied agents exist and may be found on the Internet and platforms such as television (e.g., Krämer et al., 2003; Chorianopoulos, 2006), as well as on virtual (e.g., Garau, 2003; Bailenson et al., 2006; Reidsma et al., 2007), mobile (e.g., Tomlinson et al., 2006), and desktop environments.

In many virtual worlds such as Second Life and in messenger services such as Yahoo! Messenger, users can create a virtual self. Such a self-representation, also called an avatar, graphically represents the user online. This research does not directly address such avatars. Commonalities are likely to exist between how users interact with embodied agents and how they use their avatars because they are both visual, often human-like, representations on the computer screen. However, processes involved in self-identity and self-consciousness may have a strong influence on how the user responds to avatars, for example, how involved the user feels with the avatar. This dissertation does not examine these factors, and hence the results may not be generalizable to human interaction with avatars.

Equipping applications with embodied agents is a relatively new trend in interface design. In the last two decades of the 20th century, human-computer interaction scholars questioned whether future computer applications would (and should) develop into ubiquitous, invisible, disappearing computers, or whether they would lead to natural, human-like interaction between humans and computational devices (e.g., Shneiderman and Maes, 1997). In fact, current technologies represent both of these very different views. On the one hand, the new generation of technological devices is moving from the desktop into everyday items such as refrigerators, telephones, and wearables such as glasses and bracelets. They have ‘disappeared’ from sight. They have gotten smaller, too. Apple’s iPod nano, a super slim musical device, is one example. On the other hand, many new technological devices have human-like features, such as voice interaction in car navigation systems (Nass and Brave, 2005), and embodied agents in a range of applications (not to be confused with ‘humanoid’ agents such as ELIZA (Feigenbaum and Feldman, 1963) that usually do not have a visual appearance). The idea is that users can interact more naturally and efficiently with embodied agents, a human substitute, than with machines, because humans communicate with other humans all the time (e.g., Picard, 1997). In addition, because people would often rather interact with people than with machines, embodied agents may enable people to interact with computers in an enjoyable, motivating, and engaging way (e.g., Maes in Shneiderman and Maes, 1997; Nass et al., 2006). Last, people may even interact more *honestly* with

[6]http://en.wikipedia.org/wiki/Bonzi_Buddy

embodied agents than with non-humanoid interfaces: one study found that a picture of a pair of eyes near a coffee machine where people had to leave money in a so called ‘honesty box’ tripled the amount of money collected (Bateson et al., 2006).

Of course, not all systems improve by incorporating an embodied agent. The benefit of using embodied agents is likely to depend on a variety of factors. Computer games such as Grand Theft Auto would be very empty without virtual creatures, and not interesting enough to visit. Applications used for story-telling or for teaching salsa dance are nothing without a virtual storyteller or salsa teacher. However, few people would miss Clippy if it totally disappeared from the screen. Perhaps a three-year old would. A young child may want to see the cute dog that makes such funny noises or the lovely cat that blinks at him or her from the computer screen. For small children, this is entertaining, whereas for others, it is disrupting. When people surf the web site of IKEA, they might ask the embodied agent Anna questions because she looks like a real IKEA employee. If she replies intelligently, they may seek her expertise again. Otherwise, they might ignore her the next time they have a question, and call customer service instead. The ‘framework for research on anthropomorphic interface agents’ of Catrambone et al. (2004) sorts the factors affecting user-agent interactions into three groups: features of the embodied agent, features of the user, and features of the task the user wants to perform.

Among first group, features of the embodied agent, *realism* has received the most attention in embodied agent research. The degree to which the agent resembles a real person or animal, in form and behavior, is likely to influence the user (e.g., Bailenson et al., 2006, see Chapter 2 for an elaboration). Many empirical studies have investigated form realism, by comparing realistic and unrealistic outer appearances of embodied agents. Koda and Maes (1996) found that people prefer realistic human faces over cartoon faces in terms of likeability and comfort. On the other hand, Catrambone et al. (2002) found that agent realism (lifelike versus iconic) had little effect on the perception of an embodied agent, but user perceptions were strongly influenced by the task. Despite the contradictory findings, in their overview of the empirical studies on embodied agents performed up to the year 2000, Dehn and Van Mulken (2000) concluded that the degree of realism is probably an important factor influencing likeability and other social evaluations.

Scholars have argued that besides realism, other agent features are also likely to affect the user. Just like in real-life, human responses are likely to depend on both the way an embodied agent looks (appearance aspects) as well as the things an embodied agent is capable of doing (functional aspects). Indeed, Catrambone et al. (2004) reviewed a range of relevant agent features, such as visual appearance, fidelity, expressiveness, personality, presence, role, initiative, speech quality, and other variables such as gender and competence. Further, Ruttkay et al. (2004) distinguish between aspects related to the embodiment (e.g., look, communication modalities), mental aspects (e.g., personality, emotions), and technical issues (e.g., the use of certain programming languages). Finally, Gong et al. (2006) created a categorical framework for (static) embodied agents by asking users to categorize pictures of embodied agents into meaningful groups. They identified four dimensions: humanness, graphic details and stylization, good-natured versus bad-natured, and gender. The above scholars thus identified similar dimensions (e.g., the agent's look and gender). Note that these categorizations emphasized appearance aspects more than functional aspects, or related aspects such as competence.

Note that, although embodied agents can be *designed* to smile, be rude, etcetera, they cannot actually *feel* happy or sad, they do not actually *have* a rude or extroverted personality (e.g., Konijn and Van Vugt, 2008). Users assign such qualities to the agents (just as we assign personalities to cars or other objects), but embodied agents do not actually possess them. User perceptions are based on design characteristics. For example, an embodied agent using speech may be perceived as more realistic than an embodied agent that is text-based.

The second group of factors that affect user-agent interactions are features of the user. This includes various characteristics such as gender, age, personality, cognitive style, background knowledge, capability, goal, computer experience and skills, psychological states (e.g., mood), and culture (e.g., cultural norms and believes) (e.g., Catrambone et al., 2004; Ruttkay et al., 2004). In the studies presented in this dissertation, we control for a range of user characteristics such as gender, age, and computer experience. More importantly, however, we especially focus on user *perceptions* in order to understand their responses to embodied agents (cf. Konijn and Hoorn, 2005). Thus, we investigate embodied agents from the user's perspective, which sections 1.4 and 1.5 discuss in greater detail.

The third group of factors affecting user-agent interactions are features of the task. They include the intent, domain, context, difficulty, timing and duration of the task, as well as consequences of the quality of task performance (Catrambone et al., 2004; Ruttkay et al., 2004). In this dissertation, we adopt the view that users perceive embodied agents as *tools* they may use within a task or goal context. That is, in user-agent interactions, users perceive the actions that the embodied agent can perform, and judge whether these actions are useful for their task (relevance), and whether they expect positive or negative outcomes, for example in terms of efficiency, effectiveness, or effort, when using the embodied agent in their task (valence). 1.3 and 1.5 elaborate on these ideas.

The studies mentioned above allow us to categorize the research on embodied agents. They indicate that the agent's outer appearance, as well as the agent's behavior, the user, and the task, may all influence user responses. However, to understand the psychological processes evoked during user-agent interactions, an enumeration of factors does not suffice. Because it was impossible to include every single factor in this dissertation, we therefore selected specific factors to study user responses to embodied agents systematically. We developed an *integrative model* as a conceptual framework to explain user responses to embodied agents. The model focuses on the psychological processes user-agent interaction evokes. More specifically, the model aims to explain user engagement with embodied agents, the user's intentions to use the agent again in the future, as well as user satisfaction.

Thus far the literature has lacked such a systematic integration of factors.

The building blocks of the integrative model were derived from different disciplines, mainly communication science and computer science. To understand the model it is necessary to understand the interdisciplinary approach.

1.3 Interdisciplinary approach

Interdisciplinary research exists at the boundaries of, and includes elements of, different scientific disciplines. The work this dissertation describes has been interdisciplinary from the start (being part of 'VUBIS' and the 'Network Institute' at the VU University Amsterdam), and is the result of collaborative work between two faculties: the faculty of Social Sciences, department of Communication Science, and the faculty of Exact Sciences, department of Computer Science. Not surprisingly, therefore, the research presented in this dissertation includes elements from Communication Science and elements from the field of Human-Computer-Interaction (HCI).

For a better understanding of what factors affect user responses in user-agent interactions, it seems obvious to start with insights from the field of HCI, typically concerned with the understanding of why people use certain technologies and why they ignore others. For a long time, the main focus in HCI was on computers as *tools* for task performance, just as a carpenter uses a hammer as a tool for making furniture. Microsoft's Office Assistants, for example, were designed to help us search our directories or to teach us something about computer use. The starting point for research in HCI was the belief that people evaluate the usefulness of a computer program (e.g., an embodied agent) mainly through effectiveness and efficiency considerations, during task performance. Early versions of the influential Technology Acceptance Model (TAM), developed by Davis (1989), for example, predicted computer technology acceptance and utilization based solely on the concepts of perceived usefulness and perceived ease of use. In general, aspects of utility and usability, such as learnability, efficiency, and memorability (e.g., Nielsen, 1993), gained the most attention in HCI. However, HCI scholars and designers realized that computer program use is not a purely rational, cognitive decision. Utility and usability alone could not explain the total experience of using computerized systems. Extended versions of TAM emphasized user motivation and emotion as major components in technology acceptance and use (Venkatesh, 2000). Thus it appeared necessary to go 'beyond usability' (Green and Jordan, 2001; Maxwell, 2001) to understand the total user experience of interacting with computers (cf. Preece et al., 2002; Shedroff, 2001; Forlizzi and Ford, 2000; Garrett, 2002).

Stemming from communication science, the work of Reeves and Nass (1996) has been influential in this field of interdisciplinary studies. Their work demonstrated that people often operate 'mindlessly' while interacting with computers (c.f. Langer, 1992). People apply social rules, norms, and expectations to computers, responding to them in a social way, as they would to real people. For example, impolite computers are deemed offensive, and they are not regarded as technologically deficient, but as socially incompetent. An embodied agent is not allowed to disappear from the screen without first saying goodbye to its user, because this behavior does not conform to the rules of leaving a social situation (Reeves and Nass, 1996). In addition, embodied agents suffer from many social categorizations, for example, based on ethnicity (Biernat and Vescio, 1993). Embodied agents with a similar ethnicity as the user, hence in-group members, are better liked and more trusted than out-group members (Dryer, 1999; Nass and Moon, 2000). The phenomenon that systems are seen as independent social actors is known as the Computers as Social Actors (CASA) paradigm (Nass et al., 1995, 1996; Reeves and Nass, 1996). In the mid-90s, the development of the first intelligent embodied agent, Microsoft's Bob, was a direct consequence of the Reeves and Nass studies. Bob is a primitive example of what eventually may be done to design computers that look and act like humans. More sophisticated embodied agents were designed in following years. The idea of human-like computers certainly fires our imagination. Science fiction movies such as Star Trek feature computers with gender, personality and emotion, and with whom it is easy and stress-free to talk.

People thus treat computers not only as tools with which to perform tasks, but also as communication partners with intentions, emotions, and personalities. Humans anthropomorphize their computers, meaning that they are 'described or thought of as having a human form or human attributes'[7]. This anthropomorphism affects how much individuals like a computer that is represented by an embodied agent. To go 'beyond usability', researchers began to combine models of human-computer interaction with insights from social science theories concerning human-human interactions. They drew upon person perception, interpersonal attraction, and psychological theories of emotion to understand human interaction with embodied agents (cf. Sproull et al., 1996; Reeves and Nass, 1996; Dryer, 1999; Nass and Moon, 2000; Bente et al., 2001).

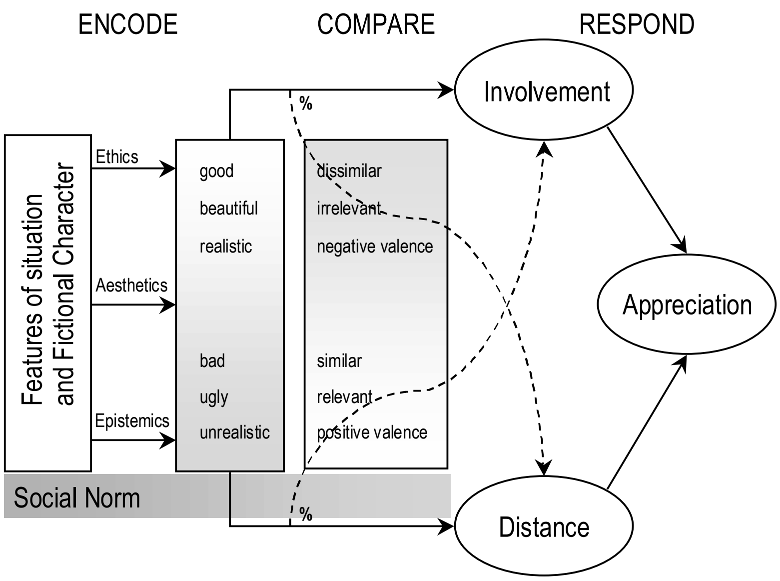
A core idea in this dissertation is that humans use their person perception skills to interpret and evaluate embodied agents. From this perspective, then the complex interplay of factors involved in human-human communication (e.g., looks, ideas, capabilities, the relevance of a person's comments and behavior in a certain context) should apply to user responses to embodied agents. Yet, since the embodied agent represents a tool, other things matter as well. In the literature, this is indicated by the various, extensive, frameworks for categorizing embodied agents. In this dissertation, this is further elaborated in the Interactive model on Perceiving and Experiencing Fictional Characters, the I-PEFiC model. The following section describes how we developed this model as a conceptual framework for studying human-agent interactions. The model integrates insights from various disciplines. This is important, because disciplines usually explore only one side of user-agent interactions, whereas a deeper understanding of this phenomenon obviously requires insights into interactions among multiple sides.

[7]Definition from Merrian-Webster online dictionary

1.4 The PEFiC model

Instead of starting from scratch to explain user responses to embodied agents (engagement, use intentions, and satisfaction), we built upon an existing model that takes a similar perspective in studying human engagement with fictional characters in literature, film, and theater (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005). This model on Perceiving and Experiencing Fictional Characters (PEFiC) is based on psychological theories of emotion, interpersonal attraction, and media entertainment (e.g., Cacioppo and Berntson, 1994; Dollard and Miller, 1950; Frijda, 1986; Lazarus, 1991). It assumes that the process of establishing engagement with film and television characters is similar to how people respond to others in real life. We similarly assume that the processes involved when humans encounter embodied agents resemble how media users respond to film and TV characters.

Figure 1.1: Perceiving and Experiencing Fictional Characters (PEFiC) (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005).



The PEFiC model itself (see Figure 1.1) as well as how the PEFiC model relates to psychological theories of emotion is described in detail in Hoorn and Konijn (2003) and in Konijn and Hoorn (2005). A short description will suffice to understand the aspects of the model that are relevant for this dissertation. PEFiC assumes that the establishment of *appreciation* or liking of of a fictional character depends on parallel tendencies to approach and avoid the character. Approach and avoidance are the backbone of the processes of *involvement* (e.g., empathy, sympathy, challenge) and *distance* (e.g., antipathy, irritation, boredom), which are the two dimensions of character engagement. In explaining engagement (involvement and distance), PEFiC assumes that when people watch a fictional character, they 'encode' that character on several dimensions[8]. Is the character good or bad (ethics)? Is the character beautiful or ugly (aesthetics)? And, is the character like a real person or a fantasy figure (epistemics or realism)? All of these factors influence how people respond to the fictional character, that is, the feelings of engagement the person has for the character. For example, as in real life (Etcoff, 1999), beauty is assumed to be more involving than ugliness.

Ethics, aesthetics, and epistemics are factors that are largely similar across (sub)cultures, within which people generally agree about what is good, what is bad, what is beautiful, etcetera. Individuals typically judge fictional characters along the ethics, aesthetics, and epistemics dimensions with reference to prevailing cultural standards, and with reference to the self only in a limited way. Responses to fictional characters may, however, also depend on viewer interpretation factors that do reference the self. Therefore, PEFiC incorporated a 'compare' phase, in which viewers 'compare' the fictional character's features with their own, such as their looks, goals, and previous experiences. The viewer determines whether the fictional character looks like him or her, and whether he behaves like the viewer does (similarity). Similarity is a key factor in interpersonal attraction, for example, for feeling sympathy and choosing friends (Byrne, 1971; Klohnen and Luo, 2003). Second, the viewer monitors whether the fictional character is more likely to harm or sustain the viewer's goals (valence), such as entertainment or learning. In the psychology of emotion, valence often reflects the direction of the affective response (positive or negative), based on the implied outcome of the encountered event (e.g., Cacioppo and Berntson, 1994). An encountered event that is about to turn out well evokes positive affect (e.g., hero helps victim). An event that is about to turn out poorly evokes negative affect (e.g., character kills hero). Valence is a key factor of engagement with fictional characters. Third, the viewer monitors whether the fictional character is important for the viewer's personal goals and concerns (relevance) (cf. Frijda, 1986; Lazarus, 1991). For example, a soap opera star is more relevant to marital problems than Clippy, but not for task execution on the computer. Relevance influences the intensity of the emotion. If the user judges a fictional character to be relevant, emotion felt towards the character will be more intense, irrespective of its positive or negative direction (valence).

PEFiC has been tested in elaborate experimental settings (e.g., Konijn and Hoorn, 2005; Konijn and Bushman, 2007). In Konijn and Hoorn (2005), engagement and appreciation were measured as a function of ethics, aesthetics, and epistemics of eight considerably different protagonists in feature movies. For example, *Gandhi* represented a good-beautiful-realistic fictional character, *Superman* a good-beautiful-unrealistic fictional character, and *Dracula* a bad-ugly-unrealistic

[8]The term 'encode' in the PEFiC model refers to the interpretation process of the viewer of a fictional character. This might be confusing, because in the literature (e.g., Zuckerman et al., 1975), the term 'decode' is also used to refer to interpretation process, whereas the term 'encode', then, refers to the transmission process.

fictional character. The experimental data generally supported the theoretical model's predictions about how fictional characters are perceived. First, the best model-fit was obtained by including all the PEFiC factors. The model-fit was sufficiently high (χ^2 = 9639.80; *df* = 4,902; *AIC* = 10755.80; *RMSEA* = 0.056), which means that, to a large extent, the PEFiC factors suffice to explain viewers appreciation of fictional characters such as those in films. Each independent variable significantly affected the dependents. For example, variation in the dimensions ethics, aesthetics, and epistemics led to variation in the intensities of involvement, distance, and appreciation.

Second, results indicated that for bad fictional characters, observers reported more distance than involvement, whereas appreciation did not strongly suffer. Thus, fictional characters can be liked in spite of distance (antipathy, cold feelings). As expected, appreciating a fictional character was best explained by *both* involvement *and* distance experiences (the explained variance of appreciation was 36%). Factor analysis showed that involvement and distance were two distinct experiences, not the ends of a continuum. They are related but can be experienced at the same time. An increasing amount of evidence suggests that involvement and distance are the two dimensions of engagement that affect character appreciation in parallel (Hoorn and Van Vugt, 2006; Konijn et al., 2007; Konijn and Bushman, 2007).

What Konijn and Hoorn (2005) and Konijn et al. (2007) found for fiction corresponds to what Priester and Petty (2001) called 'subjective ambivalence', that is, a conflict in real life between simultaneously occurring positive and negative attitudes towards an object or person. For example, a mother may feel involved with a boy because he is her son, but at the same time, she feels distanced because he has just stolen something from a store. Interestingly, feelings of ambivalence, which are assumed to cause discomfort in real life, can be used in fiction to enhance pleasure (Konijn and Hoorn, 2005).

Third, results showed that, in general, positive features (e.g., beauty) of a fictional character will produce positive effects (involvement), and negative features of a fictional character (e.g., ugliness) will produce negative effects (distance). However, the reverse may occur as well. Sometimes, the 'bad' guy (e.g., Dracula) was liked more than the 'good' guy: Some liked it bad. The direction of the effect of a dimension such as ethics may depend on the context and the viewer's goal. If you feel like being thrilled you may appreciate a character like Dracula, whereas you may appreciate Gandhi as a historical person. In addition, ugliness was more involving than beauty if the fictional character was bad, for example, a thief. Or, the distancing effect of ugliness could be compensated for by moral goodness. Thus, other features of the character may determine whether positive features lead to positive or negative effects. Konijn and Hoorn (2005) found that involvement with and distance towards a fictional character depend on a complexity of features, such as its beautiful appearance, its moral behavior, and personal relevance for the viewer.

1.5 The I-PEFiC model: Adjusting PEFiC for the embodied agent domain

Because this dissertation focuses on embodied agents, not fictional characters, we adjusted the PEFiC model for the embodied agent domain. Note that, whereas the basic ideas of PEFiC can also be found in other theories on interpersonal (human-human) attraction, we choose to use the PEFiC model as our starting point as it concerns attraction with *non*-human, or *non*-real, characters. Because of the apparent commonalities between fictional characters and embodied agents, PEFiC was a good starting-point for studying user responses to embodied agents. However, the differences between television viewing and user-agent interactions should also be considered. A main difference is that, unlike film characters, embodied agents are tools with which one can accomplish a task. Therefore, we adjusted the PEFiC model to the embodied agent domain using theories from HCI. Whereas PEFiC aims to explain viewers' engagement and final appreciation, I-PEFiC aims to explain engagement with, intentions to use, and final user satisfaction with embodied agents. The PEFiC model was based on empirical work. I-PEFiC, however, is a conceptual or theoretical model, in need of empirical testing. Our empirical studies form the body of this dissertation. Theoretically, our reasoning was as follows:

In general, people may approach embodied agents similar to the way in which they approach fictional characters. A computer user's encounter with an embodied agent triggers an engagement process. The engagement process, as the PEFiC model describes, is likely to work similarly for interaction with embodied agents. Ethics, aesthetics, and epistemics (or realism) are likely to affect interactions between humans and embodied agents. For example, more than 20,000 people downloaded 'Virtual Katja'[9] in the first week after her introduction. This agent looks like the real-life Katja, a famous Dutch television actress and presenter (realism), well-known for her attractive appearance (aesthetics). However, some time later, people accused agent Katja of carrying spyware ('bad', ethics) and stopped using her or removed her from their computers[10] (non-use).

However, there are several differences between television viewers perceiving fictional characters and computer users perceiving embodied agents. A first difference concerns how people 'stay in touch' with the characters. If traditional television viewers want to see a specific game show host or actor, they are generally dependent on the broadcasting schedule for particular shows (e.g., Vorderer et al., 2004). Computer users, however, can interact with embodied agents whenever they like, because they may access the agents at will. A second difference is that, whereas viewers often intend to satisfy entertainment needs by watching films (e.g., Vorderer et al., 2004), computer users have a larger spectrum of needs that may be satisfied by interacting with the computer, such as finding information, getting advice, or completing a task efficiently . A third difference from television

[9] http://www.katja-schuurman.com, Retrieved January 1st 2008
[10] http://www.nu.nl/news.jsp?n=514923& c=50, Retrieved January 1st 2008

viewers is that computer users *interact with* computers and embodied agents in order to perform a task or reach a goal. Computer users have some control over the course of interaction with software applications, whereas television viewers are not able to influence the course of a television program or film. Although with 'interactive television' the difference between the audiences may decrease, computer users are generally more active than television viewers. Therefore, interaction-related experiences should supplement the engagement process.

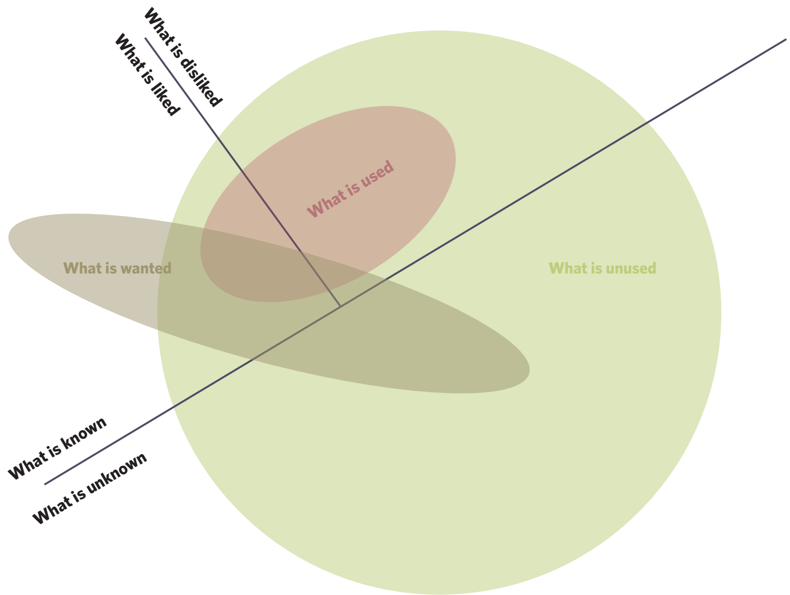
We modified the PEFiC model to incorporate the interactive nature of user-agent encounters and to explain *intentions to use* embodied agents. We used two well-known HCI theories - technology acceptance theory (Davis, 1989; Venkatesh et al., 2003, see also section 1.3) and affordance theory (among others Gibson, 1979; Gaver, 1991; McGrenere and Ho, 2000). Based on these theories, the Interactive PEFiC model (I-PEFiC, Figure 1.3) includes an 'interaction process' which focuses on the user's perception of embodied agents as *tools* that they may use within a task or goal context. Thus, whereas the engagement process concerns the character-side of user-agent interactions, the interaction process concerns the tool-side or system-side of user-agent interactions. Affordances are the possibilities for action that the software offers to the user (cf. McGrenere and Ho, 2000; Gibson, 1979), and affordance evaluations are central to the interaction process. Affordances in the virtual world operate in a way similar to the real world. First, they exist relative to agent features *and* the user (see also section 3.2.2). A window, for example, only affords opening to individuals who have hands and the strength to do so. Similarly, a virtual advisor that is text-based only affords asking advice to those computer users that can see the agent and have hands to use a keyboard.

Second, affordances have a 'quality', that is, they can be an 'aid' (positive) or an obstacle (negative). For example, it might be difficult to open a window with old or broken hinges - the affordance of opening the window can be of good or bad quality. Similarly, users may perceive the provided advice of an embodied agent to be of good or bad quality.

Third, a person may or may not want to use an affordance, depending on the situation. For example, if the temperature outside is 35 degrees Celsius, the affordance that the window can be opened becomes relevant, and the person is likely to use it. If the temperature outside is 10 degrees Celsius, the person will probably leave the window closed and opt to use the affordance of a radiator instead. Similarly, if an affordance of an embodied agent is relevant for the user, he or she is likely to use it. This is especially true if the user expects positive outcomes, for example, in terms of efficiency, effectiveness, or effort when using the embodied agent in a task (valence). If the agent seems irrelevant, the user will probably leave it untouched. The interactive side (affordances) yields intentions to use a system. Although use intentions may not *always* lead to actual usage, use intentions are strong predictors of actual usage (cf. Fishbein and Ajzen, 1975; Ajzen, 1991; Davis, 1989; Venkatesh et al., 2003). Therefore this dissertation did not need to study actual usage.

Affective responses (e.g., involvement, distance) and behavior (e.g., use intentions, actual use) often do not go hand in hand. For example, Figure 1.2, was created from data collected by talking to users of an intranet system. It shows that

Figure 1.2: Model on user responses to features of an intranet system. Most features that were used were disliked (Saffer, 2007). Reprinted with permission of Dan Saffer.



most of the features of the intranet system that were used, were also disliked. Saffer (2007, page 92) contends:

> 'The data revealed that much of the intranet wasn't being used, even features that users said they wanted. Part of the reason these features were unused was because users didn't know they existed - they were buried in the system. The data also showed that most of the features the users did know about, they disliked. Most of what the users wanted in an intranet was unused or unknown or not in the system.'

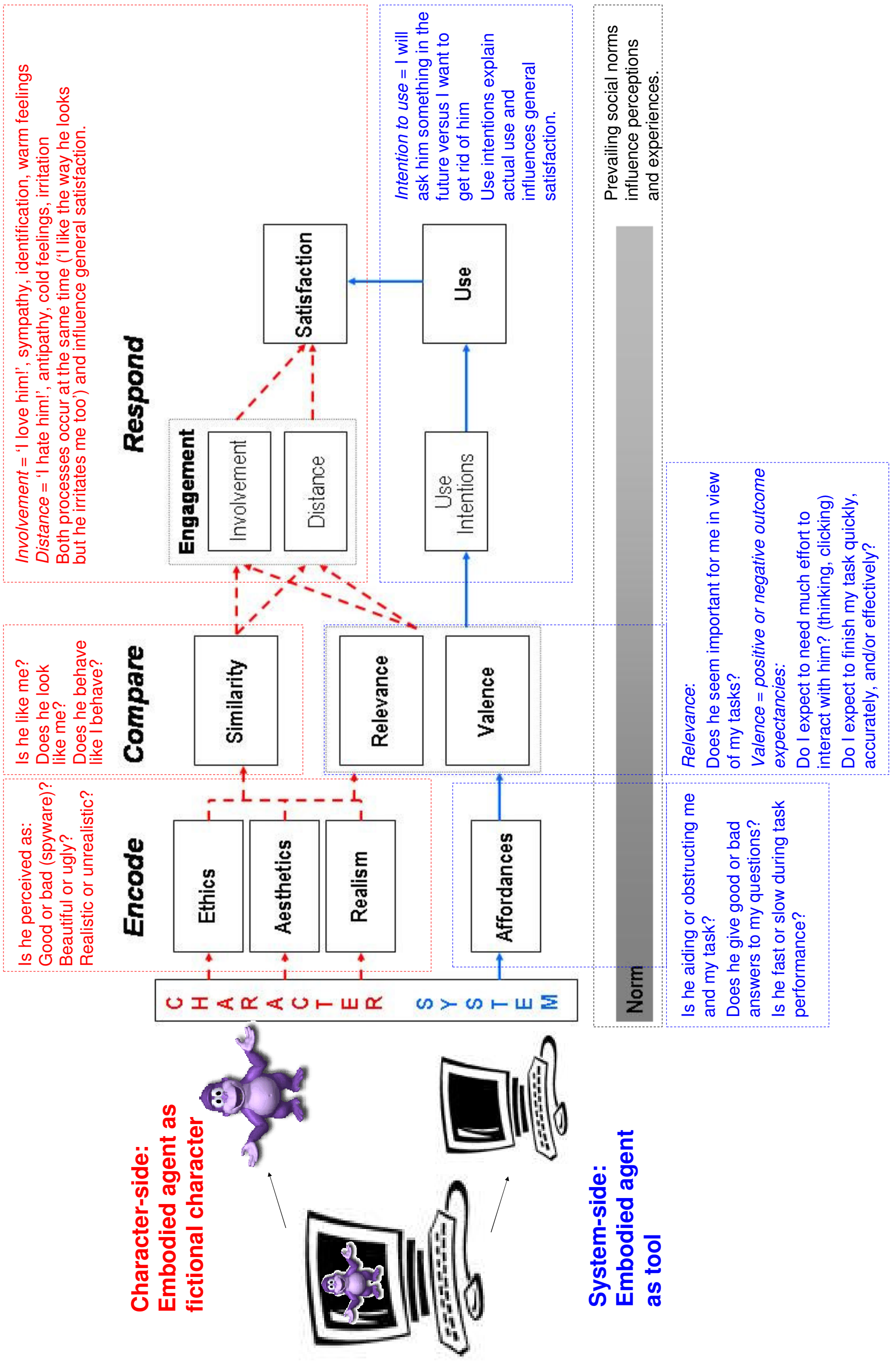
Similarly, the relationships between involvement with, distance towards, and using an embodied agent are likely to be complex. People might feel distant from an agent 'on a personal level' but still use it. Perhaps, they have no other way to perform a task (e.g., buy something online via a virtual product salesman) or someone obliges them to do so. Or, people might feel involved with an embodied agent 'on a personal level', but not intend to use it. They might not know the features of an embodied agent that may help them during task performance. Or, they might think it is inefficient to use the agent (e.g., in the case of misplaced pro-active behavior of an Office Assistant), or because they think they cannot successfully execute the behavior, that is, they might have a lack of self-efficacy (cf. Bandura, 1997; Ajzen, 1991). Therefore, we treat engagement (involvement and distance) and use intentions as separate factors that are not necessarily interconnected.

The parallel occurrence of involvement and distance in the PEFiC model results in a measure of appreciation, but in human-computer interaction, user satisfaction is seen as the more important end-goal. Therefore, the ultimate user response in I-PEFiC is user satisfaction, of which appreciation for the agent may be a part. Note that these end-states are related: satisfied users are often happy users. Following the PEFiC way of reasoning, the I-PEFiC model predicts that involvement and distance (together engagement) both influence user satisfaction. In addition, I-PEFiC predicts that satisfaction with a character not only depends on the character-side but also on the system-side of embodied agents. This argument stems from HCI literature indicating that user satisfaction is related to usability and utility aspects of a system experienced during task performance (e.g., Gelderman, 1998; Bevan and Maclead, 1994; Lindgaard and Dudek, 2003). Hence, user satisfaction is expected to depend on the engagement and interaction processes with the embodied agent.

The conceptual I-PEFiC model is quite complex because it combines theories from various disciplines into one integrative model, covering a character engagement model, affordance theory, and TAM. Basically, the Basically, the model comprises of two layers. The first layer concerns the character-side of embodied agents and describes an engagement process. The second layer concerns the tool- or system-side of embodied agents, and describes an interaction process. These layers come from different theoretical backgrounds. . It is not yet clear whether effects *cross* from one layer to another, either directly or indirectly, via perceptions of relevance and valence. For example, does the typical engagement factor realism affect use intentions? Or, does the typical interaction factor affordances affect engagement? Such '*cross-over effects*' will also be studied in the present dissertation. To better understand the full of the I-PEFiC model, Figure 1.3 as well as the glossary on page 179 provide an overview, while below we elaborate on a few important issues.

Distinction between designed and perceived features. I-PEFiC clearly distinguishes between the designed features and the perceived features. Designed features are concerned with, for example, how to give an agent an extrovert personality or a particular quality of speech. In I-PEFiC, the design features of embodied agents are part of the left box called 'character system' (see Figure 1.3). Perceived features are concerned with perceptions of a particular agent, such as the perception of beauty, ethics, and relevance. I-PEFiC includes perceived features in the 'encode' and 'compare' phases. In the experiments presented in this dissertation, the design of embodied agents (for example, a beautiful versus an ugly agent) was manipulated to affect the perception of these features, and we measured the

Figure 1.3: I-PEFiC with a short explanation of each factor.



perceptions, e.g., perceptions of beauty, evoked in users confronted with a particular embodied agent. User perceptions are often related to design characteristics, for example, an embodied agent with a small waist is generally perceived as more beautiful than an embodied agent with a large waist. Sometimes, however, variations of a certain design factor can also affect other perceptions unintentionally. A realistic agent might be perceived as more beautiful than an unrealistic agent. Therefore, we measured a range of user perceptions in each study.

User differences. Individual differences and characteristics such as personality, beliefs and attitudes, contribute to users' responses to computers (e.g., Brosnan, 1998; Ajzen, 1991; Johnson and Palmer, 2006). In addition, although some standards of beauty are universal (e.g., average face shape and symmetry, see, among others, Johnston and Oliver-Rodriguez, 1997), perceptions of beauty may differ among users and cultures (e.g., Western versus African cultures, Rguibia and Belahsen, 2006). Because such individual characteristics may also affect user responses to embodied agents, we include these in the I-PEFiC model. That is, I-PEFiC is directed at psychological variables (user perceptions) that vary among individual users because they depend on user characteristics. Obviously, compared to a male user, a female user will perceive a female agent as being more similar. Or, a funny agent may be perceived as relevant when the user wants to be entertained, and irrelevant when the user wants to finish a task efficiently. A user with computer anxiety may expect negative outcomes (negative valence) from trying to use an agent - 'the system will crash if I touch it', whereas other users have no such fear and expect positive outcomes (positive valence). Thus, I-PEFiC takes user differences into account by focusing on user perceptions.

Social norm. PEFiC states that *social norms* may affect the engagement process. In technology acceptance literature (e.g., Venkatesh et al., 2003), as well as in related theories such as the theory of planned behavior (Ajzen, 1991), social norms are regarded as a factor influencing the user. Norms are social phenomena, which group members propagate through communication (Lapinski and Rimal, 2005). What is considered as 'good', 'bad', 'beautiful', or 'ugly' depends to a large extent on the prevailing societal norms. Today, most cultures believe a slim body to be ideal, whereas four centuries ago, the round, voluptuous curves may have been considered beautiful. Although social norms are important, in order to narrow the scope of this dissertation, social norms were not investigated. For the same reason, this dissertation also does not examine other societal factors that may affect system use, such as the social position of users in society (e.g., Van Dijk, 2006).

Involvement and distance. In following PEFiC, I-PEFiC assumes that the establishment of engagement with an embodied agent consists of involvement and distance. A user might feel sympathy for an agent but simultaneously think the agent is boring (for example, a virtual news-reader). This is in line with studies that show that negative and positive affect are relatively independent factors (e.g., Diener and Emmons, 1985; Cacioppo and Berntson, 1994; Priester and Petty, 2001). We treat involvement and distance as separate factors and test whether this distinction is valid in the embodied agent domain, because researchers are still debating about whether positive and negative affect should be treated as one factor or as two separate factors (e.g., Brehm and Miron, 2006).

Positive effects of negative dimensions. I-PEFiC assumes that most often, positive dimensions (e.g., beauty) will result in positive effects (involvement, intentions to use), and negative dimensions (e.g., ugliness) in negative effects (distance, intentions not to use). However, one should not be too surprised to find positive effects of negative dimensions as well as negative effects of positive dimensions, because previous studies on fictional characters have found such effects (e.g., 'some like it bad' in Konijn and Hoorn, 2005). Perhaps ugly embodied agents (negative dimension) are still involving (raising pity). Or, the user may prefer an unrealistic agent to a realistic one, because the user is disappointed that a realistic agent does not live up to the expectations that realism evokes. User responses to embodied agents become even more complex because they feature in interactive contexts. In games, for example, obstacles during interaction promote a positive experience such as challenge or fun, and may evoke intentions to use an agent (e.g., Song and Lee, 2007). However, it is not yet clear which features affect the user under what conditions, in what direction, and to what extent. At this point, therefore, precise hypotheses about user responses to specific factors cannot be given. In the chapters to follow, we theoretically elaborated on the model and, zooming in on specific aspects of the I-PEFiC model, formulated more specific hypotheses and tested them empirically.

1.6 Structure of the dissertation

In each study presented in this dissertation, we manipulated one or several I-PEFiC factors. Table 1.1 provides an overview. In each study, we measured the factors by means of user-perception questionnaires and tested the reliability and discriminant validity of the factors. Appendix B in Chapter 2 and Appendix A in Chapter 3 show the exact items of the user perception questionnaires and the reliabilities of each scale used in the first two studies.

The dissertation incorporates four articles, dealing with effects of features of embodied agents on user responses. Although a different term (interface character) was used in two published articles, the term 'embodied agent' is used consistently throughout the dissertation, for the reader's convenience. The contents of the chapters 2 to 5 in this dissertation are identical to those of published articles or manuscripts submitted for publication. Similar to the articles, each chapter starts with an abstract and has an introduction, method section, results section, and conclusion and discussion section of its own. Parts of this introduction chapter and the discussion chapter are included in a book chapter (Van Vugt et al., 2008).

Table 1.1: Overview of manipulated factors, user perception variables, and dependent variables per chapter.

	Character- or system-side of I-PEFiC	**Manipulated factor(s)**	**User perception variables**	**Dependent variables**
Chapter 2	Character-side	(Form) Realism	Ethics, Aesthetics, Realism, Similarity	Involvement, Distance
	System-side		Relevance, Valence	Task performance Satisfaction
Chapter 3	Character-side	Aesthetics	Ethics, Aeshetics, Realism, Similarity	Involvement, Distance
	System-side	Affordances	Affordances, Relevance, Valence	Use intentions Satisfaction
Chapter 4	Character-side	(Facial) Similarity	Similarity	Involvement, Distance
	System-side	Affordances	Affordances, Relevance, Valence	Use intentions
Chapter 5	Character-side	(Body shape) Similarity, Idealness	Ethics, Idealness (a measure related to Aesthetics), Similarity	Involvement, Distance
	System-side		Relevance, Valence (not reported in this Chapter, but in Chapter 6, section 6.1)	Use intentions

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211.

Alam, M. and Dover, J. S. (2001). On beauty. evolution, psychosocial considerations, and surgical enhancement. *Archives of Dermatology*, 137:795–807.

Aylett, R. S., Paiva, A., Woods, S., Hall, L., and Zoll, C. (2007). Expressive characters in anti-bullying education. In *Animating expressive characters for social interaction*. John Benjamins. In press.

Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *PRESENCE: Teleoperators and Virtual Environments*, 15(4):359–372.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company, New York.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2:412–414.

Bente, G., Krämer, N. C., Petersen, A., and de Ruiter, J. P. (2001). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, 25(3):151–166.

Bevan, N. and Maclead, M. (1994). Usability measurement in context. *Behaviour and Information Technology*, 13:132–145.

Bickmore, T. and Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):293–327.

Biernat, M. and Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology*, 29:166–202.

Brehm, J. W. and Miron, A. M. (2006). Can the simultaneous experience of opposing emotions really occur? *Motivation and Emotion*, 30(1):13–30.

Brosnan, M. (1998). *Technophobia*. Routledge, London.

Byrne, D. (1971). *The attraction paradigm*. Academic Press, New York.

Cacioppo, J. T. and Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115:401–423.

Cash, T. F., Gillen, B., and Burns, D. S. (1977). Sexism and beautyism in personnel consultant decision making. *Journal of Applied Psychology*, 62:301–310.

Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (2000). *Embodied conversational agents*. MIT Press.

Catrambone, R., Stasko, J., and Xiao, J. (2002). Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 166–171.

Catrambone, R., Stasko, J., and Xiao, J. (2004). ECA as user interface paradigm. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 239–270, Dordrecht / Boston / London. Kluwer.

Chorianopoulos, K. (2006). Animated character likeability revisited: The case of interactive tv. *Journal of Usability Studies*, 1:171–184.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–339.

Dehn, D. M. and Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22.

Diener, E. and Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47(5):1105–1117.

Dipboye, R. L., Fromkin, H. L., and Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology*, 60:39–43.

Dollard, J. and Miller, N. E. (1950). *Personality and psychotherapy. An analysis in terms of learning, thinking, and culture*. McGraw-Hill, New York.

Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence*, 13(3):273–295.

etcoff, N. (1999). *Survival of the prettiest. The science of beauty*. Anchor Books, New York.

Feigenbaum, E. A. and Feldman, J., editors (1963). *Computers and Thought*. McGraw-Hill, New York.

Fishbein, M. and Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley, Reading, MA.

Forlizzi, J. and Ford, S. (2000). The building blocks of experience: an early framework for interaction designers. In *Proceedings of the conference on Designing interactive systems*, pages 419–423. ACM Press.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press, Cambridge.

Garau, M. (2003). *The impact of avatar fidelity on social interaction in virtual environments*. PhD thesis, University College London.

Garrett, J. J. (2002). *The elements of user experience, user-Centered design for the Web*. New Riders Publishing in conjunction with the American Institute of Graphic Arts.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 79–84. ACM Press.

Gelderman, M. (1998). The relation between user satisfaction, usage of information systems and performance. *Information and Management*, August:11–18.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gong, L., Reinhard, C. D., Lu, T., and Kerr, Z. Y. (2006). Perception and categorization of computer-generated characters. Paper presented at the annual meeting of the International Communication Association, Dresden, Germany.

Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors (2006). *Proceedings of Intelligent Virtual Agents 2006, IVA'06, Marina del Rey, CA, USA, August 21-23, 2006*, volume 4133 of *Lecture Notes in Artificial Intelligence*. Springer.

Green, W. S. and Jordan, P. W. (2001). *Pleasure with products: beyond usability*. Taylor and Francis.

Hoorn, J. F. and Konijn, E. A. (2003). Perceiving and experiencing fictional characters: An integrative account. *Japanese Psychological Research*, 45(4):221–239.

Hoorn, J. F. and Van Vugt, H. C. (2006). The role of social norm in user-engagement and appreciation of the web interface agent Bonzi Buddy. In Gratch et al. (2006), page 456.

James, H. and Edel, L. (1987). *Henry James: Selected letters*. Harvard University Press.

Johnson, R. D.and Marakas, G. M. and Palmer, J. W. (2006). Differential social attributions toward computing technology: An empirical investigation. *International Journal of Man-Machine Studies*, 64:446–460.

Johnston, V. S. and Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *Journal of Sex Research*, 34(2):188–198.

Klohnen, E. C. and Luo, S. (2003). Interpersonal attraction and personality: what is attractive–self similarity, ideal similarity, complementarity or attachment security? *Journal of Social and Personality Psychology*, 85(4):709–722.

Koda, T. and Maes, P. (1996). Agents with faces: The effect of personification of agents. In *Proceedings of the 5th IEEE International Workshop on Robot and human Communication (RO-MAN'96)*, pages 189–194.

Konijn, E. A. and Bushman, B. J. (2007). World leaders as movie characters? Perceptions of J.W. Bush, T. Blair, O. Bin Laden, and S. Hussein. *Media Psychology*, 9(1):157–177.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A., Nije Bijvank, M., and Bushman, B. (2007). I wish I were a warrior: the role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, 43(4):1038–1044.

Konijn, E. A. and Van Vugt, H. C. (2008). Emotions in mediated interpersonal communication: Toward modeling emotion in virtual humans. In *Mediated interpersonal communication*, pages 100–130. Lawrence Erlbaum Associates/ Taylor and Francis/ Routledge.

Krämer, N. C., Bente, G., and Piesk, J. (2003). The ghost in the machine. the influence of embodied conversational agents on user expectations and user behaviour in a tv/vcr application. Paper presented at the IMC Workshop 2003, Assistance, Mobility, Applications, June 2003, Rostock, Germany.

Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*, pages 289–305.

Langlois, J. H., Ritter, J. M., Roggman, L. A., and Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27:79–84.

Langlois, J. H., Roggman, L. A., and Ricser-Danner, L. A. (1990). Infants' differential social responses to attractive and unattractive faces. *Developmental Psychology*, 26:153–159.

Lapinski, M. K. and Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, 15(2):127–147.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press, New York.

Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B., and Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97*, pages 359–366, Atlanta.

Lindgaard, G. and Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15(3):429–452.

Mateas, M. and Stern, A. (2003). Faade: an experiment in building a fully-realized interactive drama. In *In game developer's conference: Game design track*.

Maxwell, K. (2001). The maturation of HCI: Moving beyond usability towards holistic interaction. In Carroll, J., editor, *Human Computer Interaction in the new Millennium*. Addison Wesley.

McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface 2000*, pages 179–186.

Moran, T. P. (1981). The command language grammar: A representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies*, 15(1):3–50.

Mueser, K. T., Grau, B. W., Sussman, S., and Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, 46(2):469–478.

Nass, C. and Brave, S. (2005). *Wired for speech: How voice activates and advances the human-Computer relationship*. MIT Press.

Nass, C., Brave, S., and Takayama, L. (2006). Human-computer interaction in management information systems: Foundations. In Zhang, P. and Galletta, D., editors, *Socializing consistency: From technical homogeneity to human epitome*, pages 373–391. M. E. Sharpe.

Nass, C., Fogg, B. J., and Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6):669–678.

Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239.

Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.

Picard, R. W. (1997). *Affective computing*. MIT Press.

Preece, J., Rogers, Y., and Sharp, H. (2002). *Interaction design: beyond human-computer interaction*. John Wiley and Sons, Inc.

Prendinger, H. and Ishizuka, M., editors (2004). *Life-like characters. Tools, affective functions, and applications*. Cognitive Technologies Series. Springer.

Priester, J. R. and Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: Interpersonal and intrapersonal antecendents of evaluative tension. *Journal of Personality and Social Psychology*, 80:19–34.

Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge University Press.

Reidsma, D., den Akker, H. J. A. O., Rienks, R. J., Poppe, R. W., Nijholt, A., Heylen, D. K. J., and Zwiers, J. (2007). Virtual meeting rooms: From observation to simulation. *AI and Society, The Journal of Human-Centred Systems, to appear*.

Rguibia, M. and Belahsen, R. (2006). Body size preferences and sociocultural influences on attitudes towards obesity among Moroccan Sahraoui women. *Body Image: An International Journal of Research*, 3(4):395–400.

Rickenberg, R. and Reeves, B. (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *CHI Letters*, 2(1):49–56.

Ruttkay, Z., Dormann, C., and Noot, H. (2004). Embodied conversational agents on a common ground. A framework for design and evaluation. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 27–66. Kluwer Academic Publishers.

Ruttkay, Z. and Pelachaud, C. (2004). *From brows to trust: Evaluating embodied conversational agents*. Kluwer Academic Publishers, Dordrecht / Boston / London.

Saffer, D. (2007). *Designing for interaction. Creating smart applications and clever devices*. New Riders.

Shedroff, N. (2001). *Experience design*. New Riders Publishing.

Shneiderman, B. and Maes, P. (1997). Direct manipulation vs. interface agents. *Interactions*, 4(6):42–61.

Song, S. and Lee, J. (2007). Key factors of heuristic evaluation for game design: Towards massively multi-player online role-playing game. *International Journal of Human-Computer Studies*, 65(8):709–723.

Sproull, L., Subramani, R., Walker, J., Kiesler, S., and Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, 11(2):97–124.

Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel, J., and Traum, D. (2006). Toward virtual humans. *AI Magazine*, 27.

Tomlinson, B., Baumer, E., and Yau, M. L. (2006). The island metaphor. In *ACM SIGGRAPH 2006 Research posters*, page 84, New York, NY, USA. ACM Press.

Van Dijk, J. (2006). *The network society*. SAGE Publications Ltd., Londen, 2nd edition.

Van Vugt, H., Hoorn, J. F., and Konijn, E. A. (2008). Modeling human-character interactions in virtual space. In Schneider, R., Jannidis, F., and Eder, J., editors, *Characters in fictional worlds*, Berlin. Springer Verlag. In press.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

Venkatesh, V. (2000). Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4):342–365.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

Vorderer, P., Klimmt, C., and Ritterfeld, U. (2004). Enjoyment: At the heart of media entertainment. *Communication Theory*, 4:388–408.

Watkins, L. and Johnston, L. (2000). Screening job applicants: The impact of physical attractiveness and application quality. *International Journal of Selection and Assessment*, 8:76–84.

Wooldridge, M. J. and Jennings, N. R. (1995). Intelligent agents: theory and practice. *Knowledge Engineering Review*, 10(2):115–152.

Zuckerman, M., Lipets, M. S., Koivumaki, J. H., and Rosenthal, R. (1975). Encoding and decoding nonverbal cues of emotion. *Journal of Personality and Social Psychology*, 32(6):1068–1076.

CHAPTER 2

Realism is not all!
User engagement with task-related embodied agents

Abstract

Human-like agents in the interface may evoke social responses in users, and literature suggests that realism is the most important factor herein. However, the effects of embodied agents on the user are not well understood. We developed an integrative framework, called I-PEFiC, to explain 'persona' and realism effects on the user. We tested an important part of the model using an experimental design in which 140 secondary school students were class-wise shown an informative virtual reality demonstration that incorporated either a realistic or an unrealistic (fantasy) embodied agent, or no agent. Findings show, first, no persona effect on task performance. We discuss how user engagement might be related to persona effects. Second, designed realism of the embodied agent contributed to user engagement when controlled for various user perceptions. Moreover, perceived aesthetics and task-relevance further influenced user engagement. Third, user engagement and task performance combined better predicted satisfaction than either one of the factors alone. In sum, several appearance- and task-related factors contributed to user engagement and user satisfaction. Thus, realism is not all.

This chapter is published as:
Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I., and Eliëns, A. (2007). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers*, 19(2):267–280.

2.1 Introduction

People can react to computers as they would react to real people (e.g., Nass et al., 1994, 1995; Reeves and Nass, 1996). An example of a user treating a human-like computer application in a social way, is: 'You are very rude! You interrupt our conversation without any real reason. I'll leave you, as you don't wish to talk to me any more.' (in De Rosis et al., 2005). Computer applications designed to resemble humans or animals have qualities, such as a human- or animal-like appearances, that are especially likely to elicit social responses. Of these computer applications, embodied agents are a special type. By embodied agents we refer to a broad range of characters that may appear on the computer interface with a life-like visual appearance. They may be computer programs themselves and possibly behave intelligently, or they may be part of a computer program (e.g., characters in games or desktop applications). They may or may not use gestures and speech to communicate and show emotions (e.g., Brave et al., 2005; Picard and Klein, 2002; Gratch and Marsella, 2005). Human-character interaction, then, is a type of human-computer interaction in which a human interacts (more or less actively) with an embodied agent.

Indeed, users respond differently to interfaces with an embodied agent than to those without an embodied agent (e.g., Takeuchi and Naito, 1995), a phenomenon known as the 'persona effect' (Lester et al., 1997). However, empirical investigations often produce conflicting results concerning persona effects on the user (Dehn and Van Mulken, 2000). Several theoretical views address the persona effect on aspects of task performance and liking. In this introduction, we outline three of these theories and then propose a new theoretical approach to studying the topic in more depth.

The first theory states that embodied agents may steer user attention, especially when they have an eye-catching appearance. This may improve task performance and liking. Some studies indeed found that an embodied agent increases a user's likeability of the system or enhances a user's learning experience (e.g., Moundridou and Virvou, 2002; Koda and Maes, 1996; Lester et al., 1997). Others found the reverse effect (e.g., Sproull et al., 1996); embodied agents may distract users from their task, which may worsen task performance and liking (e.g., Takeuchi and Naito, 1995).

Second, social presence theory explains that in real life, the mere presence of another person increases anxiety and lowers levels of performance on complex tasks, but might facilitate simple task performance (Rickenberg and Reeves, 2000). Similarly, embodied agents might evoke feelings of presence of 'another person', which might consequently affect task performance either positively or negatively, depending on task difficulty (e.g., Choi et al., 2001; Rickenberg and Reeves, 2000; Nowak and Biocca, 2003). In line with these ideas, Rickenberg and Reeves (2000) found that a 'monitoring' agent led to decreased task performance in complex tasks. On the other hand, Beun et al. (2003) found that people had better memory recall when an agent was present in the interface. In addition, some studies did not find a persona effect on task performance or recall at all. Mulken et al. (1998), for example, found that agent presence had neither a positive nor a negative effect on comprehension and recall. Likewise, no persona effects were found in a tutoring domain (Moundridou and Virvou, 2002) and in an advice giving domain (Xiao et al., 2004).

Third, from a designer's point of view, the differentiating factor between agent interfaces and traditional interfaces is human- or animal-likeness. Obviously, then, the degree to which the agent resembles a real person or animal, in form and behavior (e.g., Bailenson et al., 2006), is likely to influence the user. Form realism was subject of investigation in several studies, by comparing realistic and unrealistic outer appearances of embodied agents. Koda and Maes (1996) found that realistic human faces are preferred over cartoon faces in terms of likeability and comfort. On the other hand, Catrambone et al. (2002) found that agent realism (lifelike versus iconic) had little effect on the perception of an embodied agent, but user perceptions were strongly influenced by the task. Despite the contradictory findings, Dehn and Van Mulken (2000) conclude that realism is likely to have an important effect on users' responses. For this reason, the present study addresses realism.

Overviews of relevant factors (e.g., Ruttkay et al., 2004; Catrambone et al., 2004) indicate that factors related to the agent's outer appearance, as well as factors related to the agent's behavior, the user, and the task, all potentially explain user responses. Therefore, this study takes a more fine-grained perspective by considering a variety of factors mentioned in the literature. In this paper, we present a model as a conceptual framework to explain liking of and user engagement with embodied agents. Such a systematic integration of factors was missing in literature thus far. The model is user-centered, as it focuses on how users perceive embodied agents within a task context. The opposite is a designer-centered perspective which may focus on the design of visual life-like appearances, communication modalities, behavior and emotion modeling.

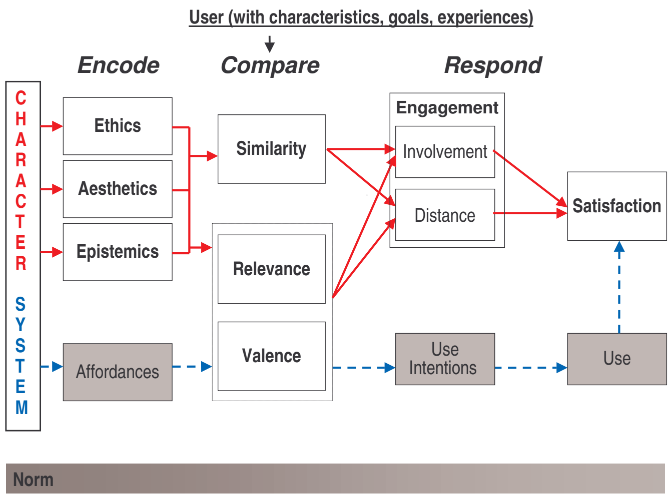
User perceptions might depend on specific design characteristics (e.g., communication modality) and task context characteristics (e.g., the user's task goal). For example, an embodied agent using speech may be perceived as more realistic than an embodied agent which is text-based. Or, a funny agent may be perceived as relevant when the user wants to be entertained, and irrelevant when the user wants to finish a task efficiently. We believe that this user-centered perspective provides us with a deeper understanding of the psychological processes underlying human-agent interactions. Further, we believe the integrative model will be useful for the HCI community, and especially those dealing with embodied agents, as it allows for a systematic empirical investigation of relevant factors, investigations into the relative importance of factors, and an integration of past research results. In addition, results can be used to inform the design of embodied agents.

Interactive PEFiC

Social science theories can be applied to study human-computer interactions, because they resemble human-human interactions (cf. *the media equation,* Nass et al., 1994, 1995; Reeves and Nass, 1996). In line with these ideas, we take a new scientific perspective on studying humans interacting with embodied agents. We apply a specific theory based on interpersonal attraction, psychological theories of emotion, and media entertainment, which explains user engagement with fictional characters, such as film and television characters (Konijn and Hoorn, 2005). This model on Perceiving and Experiencing Fictional Characters (PEFiC) (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005) integrates several factors that contribute to human engagement with fictional characters. We may conceive of embodied agents in a way similar to film characters.

Unlike film characters, however, people can actively interact with embodied agents. To account for this *interactive* nature, the PEFiC model was adjusted to include engagement during human-agent interaction, as well as embodied agent use. The model is called *Interactive PEFiC*, which is abbreviated to *I-PEFiC*, see Figure 2.1 (Van Vugt et al., 2004, 2006). In the present study, we focus on the engagement process evoked during human-agent interaction, whereas the lower part of the model (affordances, use intentions, and use) is addressed elsewhere (Van Vugt et al., 2006,b).

Figure 2.1: Graphical representation of the I-PEFiC model. Note: The grey areas and dashed lines indicate factors and processes that are not investigated in the present study (see Van Vugt et al., 2006).



In analyzing the user's experience towards an embodied agent, I-PEFiC distinguishes between three phases, *encode*, *compare* and *respond* (see Figure 2.1).

Typical factors in the encode phase of embodied agent engagement, each modeled with a positive and negative dimension, are Epistemics (realistic vs. unrealistic), Aesthetics (beautiful vs. ugly), and Ethics (good vs. bad). *Epistemics* relates to the extent to which the embodied agent is perceived as a realistic representation of a real life character. Both form and behavioral realism may contribute to perceived realism (Bailenson et al., 2006; Guadagno et al., 2007). Behavioral realism of embodied agents is concerned with, for example, the agent's facial expressions, body and head movements, gestures, eye contact and gaze (e.g., Rosis et al., 2003; Bailenson et al., 2001; Cassell et al., 1994; Cassell and Thrisson, 1999), as well as the agent's abilities, intelligence, conversational and social behavior (Hayes-Roth, 2003; Dehn and Van Mulken, 2000). Form realism is concerned with the outer appearance of the embodied agent. Does it resemble an existing living creature, whether human- or animal-like, or is it an unrealistic fantasy creature? Form realism is not likely to affect the dynamics of user interactions with an embodied agent as much as behavioral realism. Form realism can, on the other hand, be important in terms of social identity, and hence engagement with the agent. Dryer (1999), for example, contends that visual appearance, in terms of form realism, influences whether an embodied agent is perceived as agreeable (i.e. involving) or disagreeable (i.e. distancing), which is in line with our I-PEFiC model. In our study, we are interested how form realism affects engagement with the agent, while controlling for other factors.

Ethics relates to how morally good or bad the embodied agent is. For example, does the embodied agent have bad intentions when communicating (e.g., spam)? Does the embodied agent comment ironically or even sarcastically on the user's actions (cf. Agneta and Frida, see Persson, 1999)? Or, does the agent provide kind, helpful instructions (e.g., Clippy[1]? In our study we kept this variable constant across conditions. We only used agents that were helpful, and hence ethically good.

Aesthetics concerns the attractiveness of the appearance of the embodied agent (is she a beauty or an ugly agent?). There are universal standards of beauty (e.g., average face shape and symmetry; among others, Johnston and Oliver-Rodriguez, 1997), that apply to real people. Similar standards probably apply to embodied agents. Ugliness is induced by deviations from the beauty standards, for instance, embodied agents having misshapen skulls or showing signs of 'physical' decay. To avoid unintended influence from aesthetics in the present study, we attempted to keep this variable constant across conditions, and we checked it post hoc.

The comparison phase entails, first, establishing one's personal relevance and valence towards the agent, which is probably the core of establishing user engagement with embodied agents (cf. Konijn and Hoorn, 2005). These factors are intertwined with the task-context in which users find themselves during human-agent interaction. Features of the embodied agent are evaluated for their relevance (importance) to task goals of the user, and for their potential to hinder or sustain one's task goals (valence). Examples of general goals in human-agent interaction are

[1]Clippy is a product of Microsoft Inc. (http://www.microsoft.com)

entertainment, efficient task completion, or learning. In an embodied agent context, valence is concerned with questions such as ‘If I use the agent, I will have fun’ versus ‘If I use the agent, I will waste time.’ Furthermore, similarity between the embodied agent and the user also influences user engagement. Agents that are more similar to the user (e.g., in looks and behavior) are likely to be more involving than dissimilar agents. Bailenson et al. (2001) contends that users treatagents embodied with their virtual selves fundamentally different from agents embodied with virtual others. In sum, the comparison factors of relevance, valence, and similarity impinge upon user engagement with the embodied agent.

The response phase, finally, concerns engagement with an agent that consists of parallel tendencies to approach and avoid the agent. This is the backbone of the processes of *involvement* (e.g., empathy, sympathy, challenge) and *distance* (e.g., antipathy, irritation, boredom). Konijn and Hoorn (2005) and Konijn and Bushman (2007) provide evidence that liking a mediated person is best explained by *both* involvement *and* distance experiences. Thus, involvement and distance are two distinct experiences that are not the ends of a continuum; they can be experienced at the same time. A user might feel sympathy for an agent and at the same time think the agent is boring (for example, a virtual news-reader). Therefore, the I-PEFiC model predicts that *satisfaction* with the agent is fed by the simultaneously active processes of getting involved with and keeping distant from the agent.

Because the PEFiC-part of the model was originally based on non-interactive engagements with fictional characters, I-PEFiC does not take specific task contexts into account. However, interacting with embodied agents is usually directed to accomplish a certain goal through tasks, for instance, via the Internet to book a flight. It is well-known that in a task context, satisfaction is related to task performance (e.g., Gelderman, 1998). Hence, user satisfaction may not only depend on engagement with the embodied agent, but also on task performance. Furthermore, engagement might also directly affect task performance. The assumption is that increased engagement with an embodied agent improves task performance when the agent’s actions are useful for the user’s task. Increased engagement may worsen task performance when the embodied agent is distracting the user from the task. Therefore, the present study explores a variety of factors that contribute to engagement with a task-related embodied agent, satisfaction with that agent, and task performance.

In general, we expect that positive dimensions of the factors (e.g., realism, beauty) result in positive effects (e.g., involvement) and negative dimensions (e.g., unrealism, ugliness) result in negative effects (e.g., distance). However, the model also accounts for the positive effects of negative dimensions as well as negative effects of positive dimensions. This may explain why embodied agents can, for example, be ugly (negative dimension) and involving (positive effect) at the same time. For example, unexperienced users may like Clippy because it tries to be helpful, even if they find it ugly. Furthermore, Konijn and Hoorn (2005) provide evidence that both realistic and unrealistic fictional characters evoke involvement, but realistic characters more so than unrealistic ones. Similarly, both realistic and unrealistic characters evoke distance, but unrealistic characters more than realistic characters. In fact, reality judgments (realistic and unrealistic) occur in parallel, and their mixed effect further contributes to engagement. A similar pattern is expected for embodied agents. Therefore, the I-PEFiC model serves as a theoretical framework to explain potential persona and realism effects on engagement, satisfaction and task performance as studied in the present study. The model takes a user-centered perspective by focusing on user perceptions, rather than focusing on design characteristics.

2.2 Hypotheses

We designed an experiment to study the effects of 'designed realism' by taking a user-centered perspective. Using the I-PEFiC model as our theoretical framework, we focus on user perceptions, including perceived realism and perceived aesthetics, to understand how user engagement is evoked. Task performance and user satisfaction will be studied as effect variables. With this study we aim to increase the understanding of the contradictory findings found in the literature thus far.

Persona hypothesis

Although the literature is unclear about whether a 'persona effect' on task performance exists, we hypothesized affirmatively. In general, we expect positive effects on task performance in particular when the agent provides task-relevant information.

> **H1** The presence of an embodied agent positively affects task performance.

Realism hypotheses

Based on previous research, Dehn and Van Mulken (2000) stated that agent realism was likely to be an important factor influencing user experiences. Specified according to earlier findings within the context of the PEFiC model (Konijn and Hoorn, 2005; Konijn and Bushman, 2007), we formulated two hypotheses:

> **H2a** Both realistic and unrealistic agents evoke involvement, but realistic agents more so than unrealistic agents.

> **H2b** Both realistic and unrealistic agents evoke distance, but unrealistic agents more so than realistic agents.

Satisfaction hypothesis

In addition, we are interested in relations between user engagement, task performance, and user *satisfaction*. The I-PEFiC model predicts that satisfaction is a function of involvement and distance, but in a task context, satisfaction is also highly related to task performance (Gelderman, 1998). Therefore, we formulated the following hypothesis:

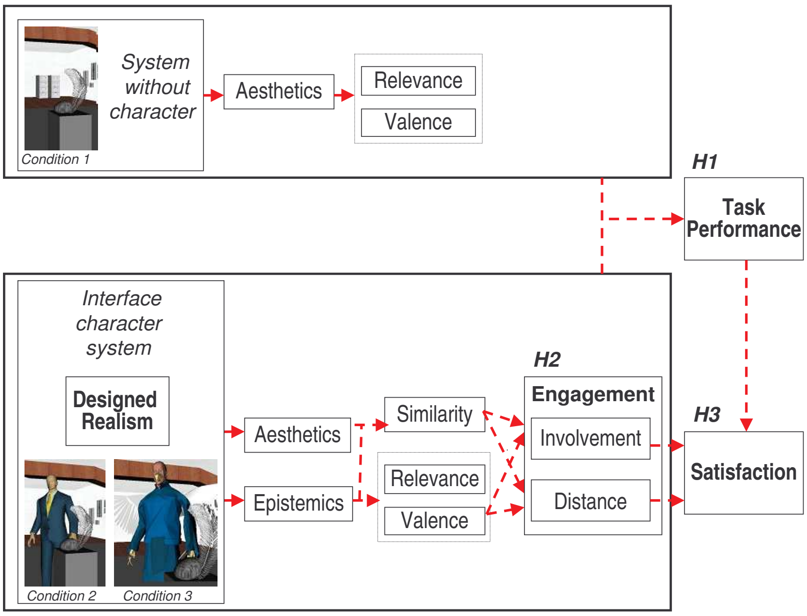
> **H3** Engagement and task performance both contribute to user satisfaction. In coalition, they better explain user satisfaction than either of the variables alone.

2.3 Method

2.3.1 Design

To address our core hypotheses adequately, we created a between-subjects design consisting of three conditions: (1) system without embodied agent (2) system with a realistic agent (3) system with an unrealistic agent. Figure 2.2 is a graphical representation of the relatively complex experimental design. The conditions 1 to 3 and hypotheses H1 to H3 are indicated in the figure. Obviously, the variables epistemics (realism), similarity, and engagement (involvement and distance) cannot be studied in condition 1 (the system without an embodied agent). As can be seen in Figure 2.2, the dependent variables task performance and satisfaction were measured in all conditions. Furthermore, the control variables perceived aesthetics, perceived (task-) relevance, and perceived valence were included in *all* conditions.

Figure 2.2: Graphical representation of the three experimental conditions in the present study.



2.3.2 Participants

The participants were adolescents (aged 12 to 18) in six classes at a Dutch middle school (N = 140). The number of participants varied from 16 to 32 participants per class, with a mean of 23. All participants used the computer at least weekly. Most participants (72%) played games on computers, boys (89%) more than girls

(56%). The number of participants and the descriptive statistics on age and gender per condition can be found in Table 2.1.

Table 2.1: Participants' characteristics per experimental condition.

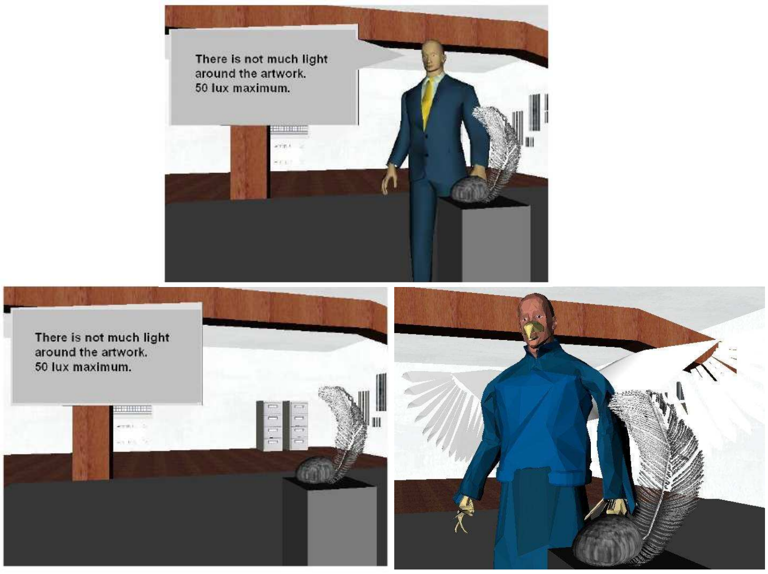
	(1) System without embodied agent	(2) Realistic agent	(3) Unrealistic agent
Number of participants	38	57	45
***M* (*SD*) of age**	16.33 (.68)	14.21 (1.69)	13.61 (.62)
Percentage boys	32%	47%	56%

2.3.3 Materials

Agent features

To avoid problems of incomparability, one should keep the various experimental conditions as similar as possible. Therefore, we designed the embodied agents ourselves, only varying the factor under study, that is, realism. We manipulated realism along the appearance dimension: the only differences between the unrealistic and realistic agent were that the unrealistic agent had a beak, wings and claws, which made it a fantasy creature (see Figure 2.3). The posture, shape, color of the clothing, and movements of the two embodied agents were equal.

Figure 2.3: The stimulus materials. The realistic agent, with text balloon, next to 'Stone and Feather' (on top). The unrealistic agent (right). The no-agent version, with text frame only (left). The text is translated from the Dutch.



Virtual environment

The stimulus materials were based on a 3D virtual environment deployed in a project performed in cooperation with the Dutch Cultural Heritage Institute (ICN)[2] in the context of the International Network for the Conservation of Contemporary Art (INCCA)[3]. In short, the idea was to develop 'digital dossiers' (Hoorn et al., 2004) for individual artworks, allowing professionals to deal with the information in an integrated, interactive fashion. We developed a prototype of a VR (the artist's studio) that contained information on the artwork 'Stone and Feather' of the Dutch artist Marinus Boezem. The environment was multimedia-enhanced; it included visual (e.g., the art work, photographs), textual (background information), and audio (e.g., audio-video recording) information. The environment allowed for the inclusion of the artist in the form of an embodied agent. The VRML based DLP+X3D platform was used to create the environment, and the DLP+X3D/STEP agent platform to create embodied agents that move naturally and use complex gestures (Eliëns et al., 2002). For more technical details, we refer to Huang et al. (2003).

Demonstration materials

For the experiment, we developed three demonstration versions (no-agent, realistic agent, unrealistic agent) of the 3D virtual environment viewable with Windows Media Player. The demonstration showed the construction of an artwork. The viewer was led through the environment while text pointed out various details of construction and the characteristics of the artwork (e.g.,'The feather is an ostrich feather'). Thus, the same communication modality (text) was used in the agent conditions as in the no-agent version to avoid that communication modality (e.g., speech versus text) might explain the effects (e.g., in studies of Sproull et al., 1996; Moundridou and Virvou, 2002). This textual information was given either as a text balloon in the embodiment conditions, or as a text frame in the no-agent condition. These balloons and frames had equal sizes (see Figure 2.3). Each version included nine text balloons.

The size of the text and the presentation speed made the demonstration easily readable from several meters away. No sound accompanied the demonstration to avoid possible confounding. The total duration of the demonstration was about 4 minutes. We used a laptop (ASUS 5600N), a projector, and a big screen (about 150 by 200 cm) to show the demonstration to the participants.

2.3.4 Procedure

At a middle school, six classes cooperated in the research that was presented to the students as 'aimed at improving computer-mediated education'. The classes

[2] http://www.icn.nl/
[3] http://www.incca.org/

participated during one day, sequentially, all in the same class room. They were instructed to watch the demonstration silently. To ensure that students paid attention to the demonstration, we told them that we would return the next day to see how well they could reconstruct the piece of art. However, as our dependent measure, the participants completed a recall test that was related to the actual reconstruction task. Directly after watching the demonstration, the participants completed this recall test as well as a user perception questionnaire (see section 2.3.5). After they finished, the participants received a small present. The next day, the participants were debriefed and thanked again for participating in the experiment. The teachers explained that the instructions concerning the reconstruction of the artwork were mainly used to guide their attention, and that it would not be necessary to actually reconstruct the artwork. Students were also told that they could contact the researchers anytime for more information concerning the artwork and the experiment.

2.3.5 Measurements

All measurements were taken by means of paper-and-pencil questionnaires. We developed a recall test to measure task performance. A structured questionnaire was used to measure user perceptions, engagement, and satisfaction.

The *recall test* consisted of 10 multiple-choice questions about the information given in the demonstration. Task performance was assessed as a percentage of correct responses. Prior to the experiment, we performed a pilot test to assess task and test difficulty. A variation in correct answers told us that the test was neither too difficult nor too simple, and that the language and questions used were comprehensible for the target test population.

The structured *user perception questionnaire* contained Likert-type scales. Each item was followed by a 6 point rating scale, ranging from 1 (do not agree at all), 2 (do not agree), 3 (barely agree), 4 (agree a little), 5 (agree), to 6 (fully agree). Items were presented in mixed order and existing scales were used wherever possible. When necessary, items were translated, adjusted to the purpose of the study, the specific material (embodied agents), and the language use of the target group of adolescents. In order to avoid directing participants in an affirmative direction (see Dillman, 2000), all scales contained indicative items (e.g.,'The figure looks natural') and counter-indicative items (e.g.,'The figure looks fake').

As an indication of how well a set of items measures a latent construct, we use Cronbach's alpha. In other words, Cronbach's alpha indicates whether a set of items is a homogeneous set that covers the meaning of the theoretical construct. The higher the Cronbach's alpha, the more reliable the generated scale is. A Cronbach's alpha above .60 generally indicates an acceptable scale (Dillman, 2000).

Perceived realism was measured with 4 items (c.f. Konijn and Hoorn, 2005). A Cronbach's alpha of .55 indicated that the statistical reliability of the perceived realism scale was marginal. Psychometric analyses gave no indicators to improve this scale. Although group-wise comparisons are still allowed, results should be interpreted with caution.

The measurement of *engagement* was based on scales used by Konijn and Hoorn (2005), distinguishing between *involvement* and *distance*. The involvement scale was measured using two items ('The figure is appealing' and 'The figure makes me happy') and was reliable according to a Cronbach's alpha of .64. The distance scale was measured using two items ('The figure is annoying' and 'The figure irritates me') and was reliable according to a Cronbach's alpha of .77.

The scales for *perceived aesthetics*, *perceived relevance*, and *perceived similarity* were also based on Konijn and Hoorn (2005). The perceived aesthetics scale consisted of two indicative items (e.g.,'The figure is beautiful') and two counter-indicative items (e.g.,'The figure is ugly'). Compared to the original scale, one item had to be discarded because of poor fit (i.e., an inter-item total correlation $<$ .30). The remaining scale was reliable (Cronbach's alpha = .77). The perceived relevance scale consisted of two indicative items (e.g.,'What the figure says, is important') and two counter-indicative items (e.g.,'If something is said in the demo, I think: never mind!'), and had a satisfactory reliability (Cronbach's alpha = .79). The perceived similarity scale comprised of only two items, one indicative item (e.g.,'I resemble the figure') and one counter-indicative item (e.g.,'The figure and I are different'). This scale was reliable (Cronbach's alpha = .82).

In addition, *perceived valence* and *satisfaction* items were based on technology acceptance theory (e.g., Davis, 1989; Venkatesh et al., 2003) and requirements analysis literature (e.g., Hoorn and Van Vugt, 2006). Perceived valence was measured using 8 items. We included various items to test whether different types of valence, concerning effectiveness, accuracy, speed and effort, would all fit into one valence-scale. The perceived valence scale consisted of four indicative items (e.g.,'I can reproduce the artwork without errors', i.e., accuracy valence) and four counter-indicative items (e.g.,'It is difficult to reproduce the artwork', i.e., effort valence). The different types of valence fit into one scale that was reliable according to a Cronbach's alpha of .90. The satisfaction scale consisted of two indicative items (e.g.,'The figure was clear') and two counter-indicative items (e.g.,'The figure was confusing'). A Cronbach's alpha of .80 indicated a reliable scale to measure satisfaction.

The scales that dealt with specific agent properties (realism, similarity, and engagement) could not be included in the questionnaire used in the no-agent condition. Where needed, items containing 'figure' were replaced by items containing 'demo'. For example, 'The demo is beautiful' (perceived aesthetics) and 'What is said in the demo, is important' (perceived relevance). The questionnaire used in the agent conditions consisted of 30 items, and the questionnaire in the no-agent condition consisted of 20 items. Additional information was collected about the students' gender, age, game usage (hours per week), computer usage (hours per week), and school level (either low or high).

Appendix A shows the recall test, translated from Dutch. Appendix B shows the items of the user perception questionnaires and the reliabilities of each scale.

2.4 Results

In this section, we first check whether the realistic agent was indeed perceived as more realistic than the unrealistic agent (manipulation check). After that, we test our hypotheses and perform additional analyses to look into user perceptions. The criterion for significance was set as alpha = .05. Note that the persona hypothesis is tested on task performance as dependent variable and the realism hypothesis on engagement (involvement and distance) as dependent variable. Finally, these dependents are tested for their combined contribution to satisfaction.

2.4.1 Manipulation check

To assess the success of our realism manipulations, an analysis of variance (ANOVA) with designed realism (Realistic versus Unrealistic) as between subjects variable and perceived realism as dependent variable was performed. Participants perceived the realistic agent as significantly more realistic (M = 3.09; SD = .98) than the unrealistic agent (M = 2.62; SD = .80): $F(1,100) = 5.90$, $p < .02$. The successful manipulation justified further hypotheses testing.

2.4.2 Testing Hypotheses

To test whether the presence of an embodied agent affected task performance (here, recall) or not (H1, persona hypothesis), we conducted a contrast analysis of variance (ANOVA). This is a solid method to compare the mean of one level (no-agent) to the mean of other levels combined (realistic and unrealistic agent). Results showed that the no-agent (M = 81% correct answers; SD = 17) and the agent (M = 78% correct answers; SD = 17) conditions led to similar levels of recall (p = .63). Thus, we did not find support for the persona hypothesis (H1) as we did not find a persona effect on task performance.

Second, to test whether the realism hypotheses (H2a and H2b) were supported by the empirical data, a multivariate analysis of variance (MANOVA) was conducted. We entered designed realism (realistic vs. unrealistic) as the between-subjects factor, and involvement and distance as the dependent variables. Results showed that the realistic and unrealistic agent evoked similar levels (p = .31) of involvement (M(realistic) = 2.11; SD = .99, and M(unrealistic) = 2.41; SD = .98) and distance (M(realistic) = 3.66; SD = 1.38, and M(unrealistic) = 3.58; SD = 1.23). Thus, we did not find empirical evidence that supported these two hypotheses. Realistic agents did not evoke more involvement than unrealistic agents (H2a), and unrealistic agents did not evoke more distance than realistic agents (H2b).

Third, we tested whether the empirical data supported the satisfaction hypothesis (H3). We expected that satisfaction using an embodied agent would be significantly influenced by task performance, involvement with, and distance toward the embodied agent. Therefore, we performed a regression analysis (method Enter), with task performance, involvement, and distance as predictors, and user satis-

Table 2.2: Means and standard deviations (SD) of perceived realism, aesthetics, relevance, valence, similarity, involvement, distance and satisfaction per experimental condition. n.a. = non applicable.

	Perc. Realism	**Perc. Aesthetics**	**Perc. Relevance**	**Perc. Valence**	**Perc. Similarity**	**Involvement**	**Distance**	**Satisfaction**
No agent	n.a.	3.12 (1.17)	2.96 (1.13)	3.52 (1.26)	n.a.	n.a.	n.a.	3.51 (1.46)
Realistic agent	3.09 (.98)	2.87 (1.07)	3.35 (1.18)	3.22 (1.07)	1.81 (1.06)	2.11 (.99)	3.66 (1.38)	3.81 (1.03)
Unrealistic agent	2.62 (.90)	2.65 (1.07)	3.87 (1.19)	3.47 (.94)	1.53 (.82)	2.41 (.98)	3.58 (1.23)	3.64 (.96)

faction as the dependent variable. The model explained 26% of the variance in satisfaction. Distance appeared to be the best predictor of satisfaction with an embodied agent (*beta* = -.41, $p < .001$, partial r = -.40, semi-partial r = -.37) followed by task performance (*beta* = .32, $p < .01$ partial r = .34, semi-partial r = .32). Involvement did not significantly contribute to satisfaction (*beta* = .07, $p = .42$). Based on these results, the satisfaction hypothesis can be partly sustained; in coalition, distance (as part of the engagement process) and task performance can better explain satisfaction than either of the variables alone.

2.4.3 Additional analyses

Thus far, we tested the effects of *designed* characteristics on users' responses. Yet, because the I-PEFiC model predicts that a number of user *perceptions* in the encoding stage are indicative to user engagement, we tested specific user perceptions in predicting user engagement. In addition, we tested the effects of user perceptions on task performance. Thus, we tested the persona and realism hypotheses again, but this time we included the user perceptions in the analyses, in addition to the persona factor and the realism factor. The participants' means on the various user perceptions, as well as on involvement and distance, are presented in Table 2.2.

First, we tested whether user perceptions could explain involvement and distance better than designed realism. Therefore, we performed an ANOVA with designed realism (realistic versus unrealistic) as the between-subjects factor, and involvement and distance as the dependents. Furthermore, we included perceived realism, perceived aesthetics, perceived similarity, perceived (task-)relevance, and perceived valence as covariates. The multivariate effects were tested using the multivariate criterion of Wilks' lamda. This statistic evaluates the multivariate hypothesis that the population means on the multiple dependent variables are equal across groups, while controlling for interdependency between dependents (e.g., Stevens, 2002). The multivariate test showed that designed realism significantly affected the dependents, that is, involvement and distance taken together (Wilks' lambda = .93, $F(2,93) = 3.47$, $p < .04$, partial η^2 = .07[4]). Furthermore, perceived

[4] Using partial eta-squared (partial η^2) as the measure of effect size. In general, the conventional

aesthetics (Wilks' lambda = .67, $F(2,93) = 32.07$, $p < .001$, partial $\eta^2 = .33$) and perceived relevance (Wilks' lambda = .79, $F(2,93) =$, $p < .001$, partial $\eta^2 = .21$) also affected the dependents. Perceived similarity and perceived valence did not significantly affect the dependents (ps $> .05$).

More specifically, further tests specified whether designed realism and the user perceptions contributed to involvement or distance, or both. First, designed realism affected involvement ($F(1,94) = 6.44$, $p < .02$, partial $\eta^2 = .06$). Of the user perceptions, only perceived aesthetics affected involvement ($F(1,94) = 19.22$, $p < .001$, partial $\eta^2 = .17$), The more beautiful the agent was perceived to be, the more involved participants were. Second, both perceived aesthetics ($F(1,94) = 25.43$, $p < .001$, partial $\eta^2 = .21$) and perceived (task-)relevance ($F(1,94) = 22.82$, $p < .001$, partial $\eta^2 = .20$) affected distance. The more beautiful and relevant the agent was perceived to be, the less distant participants felt. Note that we did not find effects of *perceived* realism on either involvement or distance. However, designed realism did affect involvement. Thus, if we control for the user perceptions in the analyses, we found that realism affected user involvement which supports realism hypothesis H2a, in contrast to results in section 2.4.2. Note that designed realism and perceived realism are highly related (we checked the manipulation of designed realism using the perceived realism scale), which makes it difficult to separate these effects. In sum, the results of this additional analysis showed that, next to designed realism, several user perceptions significantly contributed to user engagement. That is, perceived aesthetics affected both user involvement and distance towards the agent, whereas perceived relevance contributed to user distance.

Second, we tested to what extent user perceptions could explain task performance. We performed a contrast analysis within ANOVA, comparing the no-agent system with the embodied agent systems (agent presence was the between subject factor), and task performance as the dependent variable. Furthermore, we entered perceived aesthetics, perceived relevance, and perceived valence as covariates. As in the foregoing analysis (concerning H1), we found that agent presence did not affect task performance ($p = .46$). In addition, none of the user perceptions significantly influenced task performance (ps $> .05$), indicating that user perceptions (of embodied agents to which users feel distant) were not decisive for task performance.

Then, we questioned whether task performance, involvement, and distance were affected by the personal characteristics of participants (gender, age, game usage, computer usage, and school level). Therefore, we performed two additional analyses. An additional ANOVA was performed with personal characteristics as covariates to see whether these could explain task performance. We found a significant main effect of school level ($F(1, 121) = 12.61$, $p < .001$, partial $\eta^2 = .09$) on task performance. Higher qualified students performed better ($M = 82\%$ correct answers; $SD = 18$) than lower qualified students ($M = 74\%$ correct answers; SD

cutoffs are .01, .06, and .14 for small, medium, and large η^2, which are even too large for partial η^2 (Green and Salkind, 2005, p. 187)

= 14). This result follows general expectations and did not affect our theoretical assumptions.

An additional MANOVA was performed to test for effects of the personal characteristics on user involvement and distance. The multivariate test showed that only age affected (one of) the dependents (Wilks' lambda = .86, $F(2, 88) = 7.30$, $p < .001$, partial $\eta^2 = .14$). Further tests showed that age affected distance ($F(1, 89) = 14.76$, $p < .001$, partial $\eta^2 = .14$), but not involvement ($p = .26$). The elder the students were, the more distant they felt towards the embodied agents. Perhaps, the elder students might have found the demonstration childish or have a more negative attitude in general.

2.5 Conclusion and discussion

The main goal of the study was to investigate persona and realism effects on agent engagement and task performance, taking a user-centered perspective. Hereby, we focused on form realism, not behavioral realism. In our study we tested an important part of the I-PEFiC model, which aims at explaining user engagement and satisfaction with embodied agents. This model integrates a variety of factors mentioned in the literature, and focuses on how users perceive embodied agents within a task context.

First, we did not find a persona effect on task performance. The presence of an embodied agent neither increased nor decreased task performance, as compared to the no-agent system. In addition, user perceptions (of aesthetics, task relevance, etc.) did not contribute to task performance. Thus, task performance remained unaffected by the presence or absence of an embodied agent. Second, the results of hypothesis testing indicated no support for the realism hypothesis. That is, whether the embodied agent had a realistic or unrealistic design did not affect users' involvement nor distance with the agent. However, if users' perceptions of the embodied agent were included in the analyses, designed realism did affect involvement (but not distance). This supports the realism hypothesis that (designed) realism of an embodied agent affects the users' involvement with that embodied agent. We did not find an effect of perceived realism, which might be due to the relatively low reliability of this scale.

Results further indicate that several user perceptions affected involvement and distance with embodied agents. It appeared that users felt more involved with an agent when they perceived it as more beautiful. In addition, users felt less distant to an agent when they perceived it as more beautiful and more relevant for the task. This is in line with results of Van Vugt et al. (2006) in which engagement was affected by both the aesthetic looks of an embodied agent, as well as the relevance of the affordances the embodied agent offered for task completion. Last, we found support for the satisfaction hypothesis. Both task performance and distance predicted user satisfaction with an embodied agent. Involvement, however, did not significantly contribute to satisfaction. Participants felt rather distant to, and not so much involved with, both the realistic and the unrealistic agent. This might explain why distance, and not involvement, was a predictor of satisfaction. This finding shows resemblances with results from Konijn and Hoorn (2005) and Konijn and Bushman (2007). They found that distance was a better predictor of liking an agent than involvement. The result is also in line with Lindgaard and Dudek (2003) and Van Vugt et al. (2006) who found that satisfaction is a subjective sum of several user experiences with the system. Both the appearance, as well as task-related features of an embodied agent affect user satisfaction with that agent. Because end-user satisfaction is seen as an important goal in user-system interaction and design, our results show that it is important to enhance both engagement and features related to the task at hand.

The present results showed that perceived aesthetics was the most important variable that influenced user engagement with the embodied agent. The more beautiful users found the agent, the more engaged they were. This is in line with results of Van Vugt et al. (2006), in which aesthetics was also the best predictor of user engagement. The result is also in line with results from studies dealing with interpersonal attraction in real life, showing that beautiful people are generally preferred over ugly people. Furthermore, aesthetics also influences users when they are interacting with computers (e.g., Tractinsky et al., 2000). Perceived aesthetics might account for some unexpected results of previous studies on embodied agents. For example, the study of Beun et al. (2003) experimentally compared three interfaces incorporating either a realistically designed, pretty female agent, a cartoon-like agent, or no agent. Participants performed better on a memory task when the female agent was present than when the cartoon-like or no agent was present. Although these results were said to support the persona effect hypothesis on task (here, memory) performance, this was only true for the female embodied agent.

Using the insights of the present study, we may understand why interaction with the female agent, but not with the cartoon-like agent, led to higher recall than in the no-agent condition. This might have been due to her aesthetic appeal. As we learned in the present study, perceived aesthetics of an embodied agent positively contributes to user engagement. It is likely that the female agent was more attractive to the (64% male) participants than the cartoon agent, and hence more engaging (involving). Possibly, the engaged students paid more attention to the story told by the pretty interface lady than the students that used the cartoon-like or the no-agent version. As a result they had a better recall. Thus, the positive effect of the agent on task performance might have been the result of the particular agent used, that is, a beautiful one.

Furthermore, in the study of Koda and Maes (1996) human and cartoon faces were used to study (form) realism effects. They stated that the level of realism of an embodied agent (referred to as level of abstraction or humanity) changes its likeability and perceived intelligence. However, it is unclear whether perceived realism was the factor that led to effects, or whether, in line with our study, perceived aesthetics was a decisive factor. After all, the human face could have been more aesthetically pleasing than the cartoon face. This is a rather fair explanation of the reported perceived intelligence effect, as empirical studies demonstrate that individuals perceive beautiful people to be more intelligent than ugly people (e.g., Kanazawa and Kovarb, 2004).

Remarkably, embodied agent research thus far has belittled the importance of aesthetics. This was also the case for a long time in other HCI fields (see Tractinsky, 1997; Liu, 2003; Norman, 2004). We believe that embodied agent research would benefit from a recognition of not only the importance of realism, but also of the importance of aesthetics in human-agent interactions (see also Van Vugt et al., 2006). How perceptions of realism and aesthetics of embodied agents interrelate, and which factors contribute most to user engagement and satisfaction, should be investigated in future studies.

In the present study, not only perceived aesthetics but also perceived relevance of the agent to accomplish a task was a predictor of engaging with the agent. That is, when the embodied agent was perceived as more relevant to the task, the user felt less distant to the agent. Unlike realism and aesthetics, relevance is a variable that is not clearly related to the agent's appearance. However, this factor is related to the user's task goal. Is the agent relevant for the user to perform the task in the particular environment? The importance of taking into account the task context in embodied agent research has been recognized (e.g., in Dehn and Van Mulken, 2000; Catrambone et al., 2002). Based on the results of our experiment within the I-PEFiC framework, we assume that it is not so much the task context itself, but the *perceived* relevance of the embodied agent in view of the task to be accomplished, that determines the users' responses. In a game, for example, the presence of an embodied agent that behaves realistically might be important to make the game entertaining (the user's goal). However, in an editing task, for example, such an embodied agent might be irrelevant for or even obstructing efficient task completion (the user's goal). When taking into account the (perceived) relevance of embodied agents in various task contexts, we might better understand why users react differently to the same agent in different task contexts.

Given that various perception variables influence user engagement with embodied agents, we might comprehend why some studies found a persona effect and other studies did not. Each study reported in the introduction of the present paper used different agents with different appearances and characteristics. In addition, the studies used different tasks of varying complexity. Thus, in the various studies, users might have perceived the embodied agents differently in terms of aesthetics, realism, and relevance. Whereas a beautiful embodied agent providing relevant information for the user's task at hand may evoke strong user engagement, an ugly embodied agent not helping the user to perform his task may evoke no engagement at all. As a result, users might have been more engaged with an agent in one study than in another study. Engaged users are likely to pay more attention to the agent and experience more social presence (e.g., Choi et al., 2001). Thus, task performance and learning may increase, and persona effects can be found. Engaging embodied agents might be advantageous in, for example, computer-mediated learning environments. In contrast, agents with which the user does not feel engaged, do not boost the attention or feelings of social presence of users, and hence, no persona effects are involved (such as in the present study). To study whether engagement is indeed a key factor in causing a persona effect, various levels of engagement with an embodied agent can be created. Then, we expect a persona effect when engagement is high, and no persona effect when engagement is low.

Ideally, stimuli differ only on *the one or a few* dimensions that are studied in a particular experimental setting, in order to attribute effects to that dimension(s). But, even when stimuli are designed to differ only in terms of realism, one cannot prevent other factors from playing a role as well. The reason is that users perceive agents on more than one dimension. These perceptions might suppress effects of the factor under investigation, or account for the effects themselves. Thus, it is important to control for factors that are likely to covary with factors under study, as we did in the present project. Overviews of relevant factors as well as our integrative I-PEFiC model can further inform future research.

Another point of methodological consideration is that our findings address a particular group of adolescents. We used a limited sample comprised of students of only one school and of a specific age range. Adolescents might care more about outer appearance in terms of aesthetics than adults do. This might partly explain the effects of aesthetics we found in the study. The participants also had a certain amount of computer experience and are used to playing games regularly. This might have biased their perceptions of the embodied agent. Often, students are every-day users of top-of-the-art games featuring life-like agents in graphically rich environments having a high degree of behavioral realism. They might have found the agent in the demonstration boring and dull, which would explain the high level of distance they felt towards the agent. On the other hand, users with limited computer and game experience might find the agent such as presented in the demonstration interesting.

Perceptions of realism are likely to be affected by both form and behavioral realism. We may question whether perceived realism will have a larger contribution to user engagement when behavioral realism is manipulated. In addition, we may question whether perceptions of aesthetics and task-relevance continue to be the main predictors of user engagement in other contexts. Further, we may question the generalizability of the results to real-life situations in which people actively interact with the embodied agent (e.g., Clippy or game characters). Then, users will experience the usability of the agent software which concerns, for example, conversational capabilities of the agent. This might not only increase the behavioral realism of the agent, but also alter the outcome expectations (valence) as formulated in the I-PEFiC model. Also, the proactive and reactive behavior of an agent might influence user responses in human-agent interactions (c.f. Xiao et al., 2003), as such a behavior might directly contribute to perceptions of realism, ethics, affordances, and valence.

In the present study, the I-PEFiC model provided a useful theoretical framework to determine factors that explain user responses to embodied agents. Mean-while, the I-PEFiC model has been tested using various embodied agents in different task contexts. For example, we focused on situations in which users actively interacted with an embodied agent (hence, studying the lower part of the I-PEFiC model which concerns the agent's affordances, see Figure 2.1). In one study we investigated the role of aesthetics and affordances in the engagement process. Therefore, we designed beautiful and ugly looking embodied agents within the Sims2 game environment with which users could actively interact (Van Vugt et al., 2006). In addition, we designed a web-based application in which users interacted with an embodied agent in the role of a health advisor (Van Vugt et al., 2006b). With our studies, we aim to provide a more comprehensive understanding of human-agent interactions.

In sum, understanding what determines user engagement is important because engagement highly predicts user satisfaction in human-agent interaction. The present study showed that realism is not all that matters for user engagement with embodied agents. User responses are also affected by the agent's aesthetic appearance and its relevance to the task. The debate on the existence or non-existence of a persona effect may still be undetermined, but we now may suspect a more complicated picture in which various factors work together in predicting engagement, task performance and user satisfaction. An embodied agent is more likely to result in persona effects when it is engaging.

2.6 Acknowledgments

We are thankful to the Faculty of Social Sciences and the Faculty of Sciences, Vrije Universiteit in Amsterdam, who granted a 4-year research track (VUBIS) to the first author. We are grateful to Gerrit van der Veer and Jan Kleinnijenhuis for their generous support and fruitful discussions in the early stages of this project. Further, we would like to thank the students of the Multimedia Casus 2004 (spring) project and especially Ricardo Tangali for constructing the stimulus materials, and students of the Blaucapel College in Utrecht (The Netherlands) for their participation. Finally, we thank Herre van Oostendorp for his insightful comments on an earlier draft of this paper.

Appendix A Recall test

The recall test is translated from Dutch.

Question 1. The feather is from a
a. ostrich b. peacock c. pheasant d. heron

Question 2. Which color does the feather have?
a. white b. grey c. light yellow d. light braun

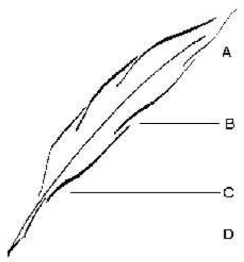
Question 3. The length of the feather is
a. seventy centimetres b. fifty centimetres c. fourty centimetres d. twenty centimetres

Question 4. The stone is a
a. marble stone b. glacier stone c. erretic boulder d. basalt stone

Question 5. The height of the stone is
a. 13.5 cm b. 16.5 cm c. 9.5 cm d. 5.5 cm

Question 6. The artwork has to meet the following requirement:
a. the feather is being kept in place by the weight of the stone
b. the end of the feather is bent around the stone
c. the stone is placed on the feather with it's flat side
d. the feather is bent after the placement of the stone

Question 7. Where is the stone placed on the feather?



Question 8. In the formation of the artwork the following requirement has to be met
a. the feather should not reach beyond the column
b. the feather has to reach beyond the column
c. the feather has to be positioned as straight as possible
d. the feather has to lie as flat as possible

Question 9. In the formation of the artwork, the following needs should be paid attention to:
a. the amount of light in the space around the artwork
b. the color of the light in the space around the artwork
c. the angle under which the light is falling at the artwork
d. the concentration of the light on the artwork

Question 10.
The column is
a. rectangular b. hexagonal c. triangular d. cube-shaped

Appendix B User perception questionnaire

Table 2.3: The reliability of each scale and the items of the user perception questionnaires. The item printed in *italics* were removed from the scale due to a bad fit.

Scale	**Reliability** (Cronbach's alpha)	**Measures for no-agent condition**	**Measures for embodied agent conditions**
Realism	.55	n.a.	The figure looks natural
			The figure looks real
			The figure is a fantasy creature
			The figure looks fake
Aesthetics	.77	The demo is beautiful	The figure is beautiful
		The demo has a professional look	The figure has a professional look
		The demo is ugly	The figure is ugly
		The demo has an amateur look	*The figure has an amateur look*
Similarity	.82	n.a.	I resemble the figure
			The figure and I are different
Relevance	.79	What is said in the demo, is important	What the figure says, is important
		I have learned something from the demo	I have learned something from the figure
		I don't care what is said in the demo	I don't care what the figure says
		If something is said in the demo, I think: never mind!	If the figure speaks, I think: never mind!
Valence	.90	I can reproduce the art work	I can reproduce the art work
		I can reproduce the art work quickly	I can reproduce the art work quickly
		I can reproduce the art work without errors	I can reproduce the art work without errors
		It is easy to reproduce the art work	It is easy to reproduce the art work
		The reproduction of the art work will fail	The reproduction of the art work will fail
		The reproduction of the art work will take a long time	The reproduction of the art work will take a long time
		I think I will reproduce the art work wrongly	I think I will reproduce the art work wrongly
		It is difficult to reproduce the art work	It is difficult to reproduce the art work
Involvement	.64	n.a.	The figure is appealing
			The figure makes me happy
Distance	.78	n.a.	The figure is annoying
			The figure irritates me
Satisfaction	.80	The demo helped me to understand the information	The figure helped me to understand the information
		The demo was clear	The figure was clear
		The demo gave exactly enough information	The figure gave exactly enough information
		The demo was confusing	The figure was confusing

Bibliography

Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., and Weisbuch, M. (2001). Intelligents agents who wear your face: User's reactions to the virtual self. In A. de Antonio, R. Aylett, D. B., editor, *Intelligent virtual agents*, volume 2190 of *Lecture Notes in Artificial Intelligence*, pages 86–99. Springer.

Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *PRESENCE: Teleoperators and Virtual Environments*, 15(4):359–372.

Beun, R.-J., de Vos, E., and Witteman, C. (2003). Embodied conversational agents: Effects on memory performance and anthropomorphisation. In Rist et al. (2003), pages 315–319.

Brave, S., Nass, C., and Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62(2):161–178.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH 94*, pages 413–420. ACM Press.

Cassell, J. and Thrisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5):519–538.

Catrambone, R., Stasko, J., and Xiao, J. (2002). Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research.

In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 166–171.

Catrambone, R., Stasko, J., and Xiao, J. (2004). ECA as user interface paradigm. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 239–270, Dordrecht / Boston / London. Kluwer.

Choi, Y. K., Miracle, G. E., and Biocca, F. (2001). The effects of anthropomorphic agents on advertising effectiveness and the mediating role of presence. *Journal of Interactive Advertising*, 2(1).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–339.

De Rosis, F., Cavalluzzi, A., Mazzotta, I., and Novelli, N. (2005). Can embodied conversational agents induce empathy in uers? In *AISB'05 Virtual Social Characters Symposium*, Hatfield.

Dehn, D. M. and Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. John Wiley and Sons, New York, 2th edition.

Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence*, 13(3):273–295.

Eliëns, A., Huang, Z., and Visser, C. (2002). A platform for embodied conversational agents based on distributed logic programming. In AAMAS Workshop - Embodied conversational agents - let's specify and evaluate them.

Gelderman, M. (1998). The relation between user satisfaction, usage of information systems and performance. *Information and Management*, August:11–18.

Gratch, J. and Marsella, S. (2005). Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, 19(3-4):215–233.

Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors (2006). *Proceedings of Intelligent Virtual Agents 2006, IVA'06, Marina del Rey, CA, USA, August 21-23, 2006*, volume 4133 of *Lecture Notes in Artificial Intelligence*. Springer.

Green, S. and Salkind, N. (2005). *Using SPSS for Windows and Macintosh. Analyzing and understanding data*. Prentice-Hall, Upper Saddle River, NJ, 4rd edition.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. (2007). Virtual humans and persuasion: the effects of agency and behavioural realism. *Media Psychology*, 10:1–22.

Hayes-Roth, B. (2003). What makes characters seem life-like? In *Life-like characters. Tools, affective functions and applications*, pages 447–462. Springer.

Hoorn, J. F., Eliëns, A., Huang, Z., Van Vugt, H. C., Konijn, E. A., and Visser, C. (2004). Agents with character: Evaluation of empathic agents in digital dossiers. In AAMAS '04 Workshop on Empathic Agents.

Hoorn, J. F. and Konijn, E. A. (2003). Perceiving and experiencing fictional characters: An integrative account. *Japanese Psychological Research*, 45(4):221–239.

Hoorn, J. F. and Van Vugt, H. C. (2006). The role of social norm in user-engagement and appreciation of the web interface agent bonzi buddy. In Gratch et al. (2006), page 456.

Huang, Z., Eliëns, A., and Visser, C. (2003). 'is it within my reach?' - an agents perspective. In Rist et al. (2003), pages 150–158.

Johnston, V. S. and Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *Journal of Sex Research*, 34(2):188–198.

Kanazawa, S. and Kovarb, J. L. (2004). Why beautiful people are more intelligent. *Intelligence*, 32(3):227–243.

Koda, T. and Maes, P. (1996). Agents with faces: The effect of personification of agents. In *Proceedings of HCI'96*, pages 98–103.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A. and Van Vugt, H. C. (2007). Emotions in mediated interpersonal communication: Toward modeling emotion in virtual humans. In Konijn, E. A., Utz, S., Tanis, M., and Barnes, S., editors, *Mediated interpersonal communication*, pages 100–130. Lawrence Erlbaum Associates/ Taylor and Francis/ Routledge.

Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B., and Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97*, pages 359–366, Atlanta.

Lindgaard, G. and Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15(3):429–452.

Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. *Ergonomics*, 46(13/14):1273 1292.

Moundridou, M. and Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning*, 18(3):253–261.

Mulken, S. V., André;, E., and Müller, J. (1998). The persona effect: How substantial is it? In *HCI '98: Proceedings of HCI on People and Computers XIII*, pages 53–66, London, UK. Springer-Verlag.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM Press.

Norman, D. A. (2004). *Emotional design. Why we love (or hate) everyday things.* Basic Books, New York.

Nowak, K. and Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494.

Persson, P. (1999). Agneta & frida: A narrative experience of the web? AAAI Fall Symposium on Narrative Intelligence.

Picard, R. W. and Klein, J. (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with Computers*, 14:141–169.

Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge University Press.

Rickenberg, R. and Reeves, B. (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *CHI Letters*, 2(1):49–56.

Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors (2003). *Intelligent Agents, 4th International Workshop, IVA 2003, Kloster Irsee, Germany, September 15-17, 2003, Proceedings*, volume 2792 of *Lecture Notes in Computer Science*. Springer.

Rosis, F. D., Pelachaud, C., Poggi, I., Carofiglio, V., and Carolis, B. D. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59(1-2):81–118.

Ruttkay, Z., Dormann, C., and Noot, H. (2004). Embodied conversational agents on a common ground. a framework for design and evaluation. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 27–66. Kluwer Academic Publishers.

Sproull, L., Subramani, R., Walker, J., Kiesler, S., and Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, 11(2):97–124.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Mahwah, NJ.

Takeuchi, A. and Naito, T. (1995). Situated facial displays: towards social interaction. In Katz, I., Mack, R., Marks, L., Rosson, M. B., and J. Nielsen, E., editors, *Human Factors in Computing Systems: CHI'95 Conference Proceedings*, pages 450–455, New York. ACM Press.

Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 115–122. ACM Press.

Tractinsky, N., Shoval-Katz, A., and Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13(2):127–145.

Van Vugt, H. C., Hoorn, J. F., and Konijn, E. A. (2004). Adapting empathic agents to user experiences. In AAMAS '04 Workshop on Empathic Agents.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and De Bie Dimitriadou, A. (2006a). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. (2006b). Why fat interface characters are better e-health advisors. In Gratch et al. (2006), pages 315–319.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

Xiao, J., Catrambone, R., and Stasko, J. (2003). Be quiet? evaluating proactive and reactive user interface assistants. In *Proceedings of INTERACT '03*, pages 383–390, Zurich, Switzerland.

Xiao, J., Stasko, J., and Catrambone, R. (2004). An empirical study of the effect of agent competence on user performance and perception. In *Proceedings of AAMAS'04*, pages 178–185.

CHAPTER 3

Affective affordances: Improving embodied agent engagement through interaction

Abstract

The nature of humans interacting with embodied agents is not well understood. The I-PEFiC model provides an integrative perspective on human-agent interaction, assuming that the processes of engagement and user interaction exchange information in explaining user responses with embodied agents. An experiment using the Sims2-game was conducted to test the effects of aesthetics (beautiful vs. ugly, as engagement factor) and affordances (help vs. obstacle, as interaction factor) of embodied agents on use intentions, user engagement, and user satisfaction. Results of the experiment showed that (1) people tended to use helpful agents more than obstructing agents, (2) user engagement was enhanced by beauty and perceived affordance of the agent whereas (3) intentions to use the agent were not affected by good looks, and (4) the most satisfied users were those that were engaged with the agent as well as willing to use it. This stresses the importance of enhancing affordances so to increase user engagement with embodied agents. The I-PEFiC model provided a valuable framework to study the (interdependent) effects of relevant factors in human-agent interaction.

This chapter is published as:
Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

3.1 Introduction

'For so long, the computer gaming industry was concerned with making their products look and sound better - and for good reason. There have been many instructional books and articles that have paced the advances in graphics and sound technology' (Mark, 2003)[1]. Yet, Mark also signals the need in the game development community to add 'true interactive gameplay into their products' (ibid.). Whereas the focus used to be on graphic design, game AI 'has only recently come into the primary focus as the next frontier of game advancement' (ibid.).

That the industry is switching from graphic to AI design may be for good reason as well, as suggested by the results of our experiment with the Sims2 game. If the gameplay was improved, not only the user's willingness to interact with the agent increased but also the user's engagement with the agent. By contrast, just improving the appearance of the agent had no effect on the willingness to interact with it. In the latest generations of computer games, hence, game AI seems to get more emphasis than graphic design when it comes to enhancing the overall user experience. The present study aims to investigate how graphic design in coalition with gameplay (i.e. time efficiency) of agents affect use intentions and engagement.

In recent years, computer programs are increasingly anthropomorphized, that is, mimicking humans in appearance, behavior, emotion (e.g., Brave et al., 2005; Picard and Klein, 2002; Gratch and Marsella, 2005) and/or emulating human communication skills. Such anthropomorphized communication partners, or *embodied agents*, feature in, for example, educational software, the Internet, games, and standard desktop applications. Although embodied agents are increasingly prevalent, it remains unclear how users decide whether or not to interact with them.

Understandably, many studies have investigated the effects of embodied agents on the user (for overviews see Dehn and Van Mulken, 2000; Ruttkay and Pelachaud, 2004). Many studies focused on realism effects by comparing realistic and unrealistic outer appearances of embodied agents. From a designer's point of view, embodied agent interfaces differ from traditional interfaces because they have human- or animal-likeness. Obviously, the degree to which the agent resembles a real person or living creature is likely to influence the user's experience and behavior in human-agent communication (e.g., Berry et al., 2005). Next to realism, other factors are also likely to influence human-agent interaction, as overviews of relevant factors show (e.g., Ruttkay et al., 2004; Catrambone et al., 2004). Despite such important insights, there is a need for an *integrative model* that takes into account the (interdependent) effects of a variety of factors that may explain human-agent interaction. Therefore, in the present study we present a model as a conceptual framework to explain user engagement with and intentions to use embodied agents. Then, we test hypotheses derived from the model. We believe such a model will be

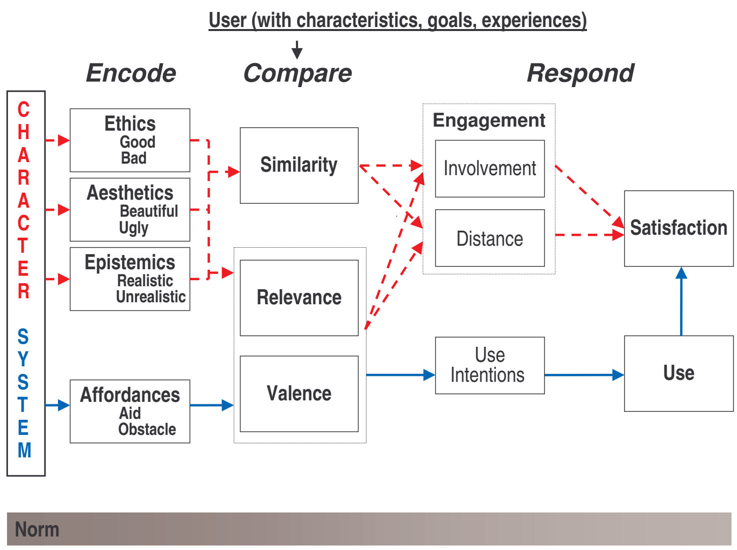
[1]Dave Mark is the president and lead designer of Intrinsic Algorithm, a computer game development company in Omaha, NE. The quote was drawn from his review of the book AI Game Development (Champandard, 2003).

useful for the embodied agent community, as it allows for (1) a systematic empirical investigation of relevant factors, (2) investigations into the relative importance of factors, (3) an integration of past research results, (4) a deeper understanding of the psychological processes underlying human-agent interaction, and (5) informing the design of embodied agents.

3.2 Conceptual framework: I-PEFiC

Literature speaks of the possibilities and advantages of applying social science theories to study human-computer interactions, because they resemble human-human interactions (cf. *the media equation,* Nass et al., 1994, 1995, 1996; Reeves and Nass, 1996). In line with these ideas, we take a new scientific perspective on studying humans communicating with embodied agents. We apply a specific theory, based on interpersonal communication theory, which explains user perceptions of film and television characters. This model integrates different factors that contribute to affective bonding between humans and mediated characters. Unlike with film and television characters, however, people can actively interact with embodied agents, thereby influencing the experience. Therefore, we adjusted the model and developed the Interactive PEFiC (I-PEFiC) model that aims to explain both affective bonding with and use of embodied agents (Van Vugt et al., 2004). This model (see Figure 3.1) integrates two main processes that are evoked during an encounter with an embodied agent: the engagement process (dashed) and the interaction process (drawn arrows).

Figure 3.1: Graphical representation of the I-PEFiC model.



In the following sections, we explain both the engagement and the interaction process in more detail. The present study focuses on the interdependencies between the two processes, which have been studied in different scientific areas.

3.2.1 The engagement process

The engagement part of the model is based on the model of Perceiving and Experiencing Fictional Characters (PEFiC) (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005). The PEFiC model is based on psychological theories of emotion and interpersonal attraction. It explains user engagement (i.e., involvement and distance (Konijn and Hoorn, 2005)) towards fictional characters, such as those from film and television.

In analyzing users' experiences towards fictional characters, PEFiC distinguishes between three phases, *encoding*, *comparison* and *response* (see Figure 3.1). Typical factors in the encode phase of character engagement, each modeled with a positive and negative dimension, are ethics (good vs. bad), aesthetics (beautiful vs. ugly), and epistemics (realistic vs. unrealistic). Comparison entails establishing personal relevance and valence towards the character. Also, Similarity between the fictitious character and the self influences user response. Finally, the response phase concerns engagement with a character, which consists of parallel tendencies to approach and avoid the character, the backbone of the processes of involvement and distance. Konijn and Hoorn (2005) and Konijn and Bushman (2007) provide evidence that the (dis)liking of a mediated person is best explained by *both* involvement *and* distance experiences. Thus, involvement and distance are distinct experiences that do not comprise two ends of a single dimension; both can be experienced at the same time. PEFiC states that the trade-off between involvement and distance better explains (dis)liking a character than either involvement or distance alone.

The engagement process is likely to work similarly for interaction with embodied agents. Because embodied agents are a special type of fictional character, we believe that epistemics, ethics, and aesthetics also influence interactions between humans and embodied agents. *Epistemics* then relates to the realistic and unrealistic outer appearances of embodied agents. Does the embodied agent resemble a living creature, whether human- or animal-like, or is it a strange, fantasy creature? Many cues might attribute to the perception of realism, for example, the agent's facial expressions, body and head movements, gestures, eye contact and gaze (e.g., Rosis et al., 2003; Bailenson et al., 2001; Cassell et al., 1994; Cassell and Thrisson, 1999), as well as the agent's abilities, intelligence, and conversational and social behavior (Hayes-Roth, 2003; Dehn and Van Mulken, 2000). *Ethics* relates to how morally good or bad the embodied agent is. For example, does the embodied agent have mean intentions when communicating (e.g., spam)? Does the embodied agent make unfriendly or even sarcastic comments on the user's activities (cf. Agneta and Frida, see Persson, 1999)? Or, does the agent provide kind, helpful instructions (e.g., Clippit)? *Aesthetics* concerns the appearance of the embodied agent (is she beautiful or ugly?). Universal standards of beauty (e.g., average face shape and symmetry, see, among others, Johnston and Oliver-Rodriguez, 1997) exist, which apply to fantasy agents as well as to real people. Deviations from the beauty standards induce ugliness when, for instance, embodied agents have misshapen skulls or show signs of 'physical' decay. Similarity, relevance, and valence judgments are based on these three 'encode' factors and impinge upon user engagement (involvement and distance towards the embodied agent). These three factors refer to user characteristics and goals. For example, is the agent similar to the user in some way (e.g., both are female), does the agent seem important (relevant) to the user's goals (e.g., efficient task completion, entertainment, or learning), and does the agent seem to harm or sustain those goals (valence) (c.f. Frijda, 1988)?

Thus, the process of establishing affective bonds between humans and embodied agents bears resemblance to how media users respond to film characters. Therefore, we began our investigation of human-agent interaction with the PEFiC model, and modified it to incorporate it's interactive nature.

3.2.2 The interaction process

The Interaction part of the model was based on affordance theory (among others Gibson, 1979; Gaver, 1991; McGrenere and Ho, 2000) and technology acceptance theories (e.g., Davis, 1989; Venkatesh et al., 2003). The interactive side (i.e., affordances or action possibilities), yields intentions to use a system or not, which appeared strong predictors of actual use (Davis, 1989; Venkatesh et al., 2003). Thus, the interaction process focuses on the user's decision of whether or not to interact with an embodied agent. The perception of affordances is typical for the interaction process.

Affordance theory

The ecological psychologist Gibson was the first to frame affordances as unified relations between the environment and an actor. 'The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill' (Gibson, 1979, p.127). Affordances can be explained as action possibilities that actors have in the environment. That is, an affordance exists relative to 1) properties of the environment and 2) the action capabilities of an actor (McGrenere and Ho, 2000). For example, a chair has the affordance of 'sitting', because of its shape, height and carrying capacity. Humans' ability to sit, the length of their legs, and their weight, enables them to sit on the chair. The concept of affordances is of particular interest in the field of human-computer interaction, which is primarily concerned with studying how properties of computers (the environment) and humans (actors) influence their interaction with each other.

Goals are central in affordance evaluations. It is important to understand that an affordance does *not change* as the needs and goals of the person change (McGrenere and Ho, 2000, interpreting Gibson, 1979). A chair affords sitting, independent of whether a person wants to sit or not. However, people's actions *do* depend on the goal context. People typically act within the environment (they use an affordance) because of a goal they want to achieve (for example, performing a task or having fun). When humans interact with computer systems (such as embodied agents), they perceive or 'encode' them *in terms of* action possibilities for goal achievement.

Affordances in I-PEFiC

Affordance theory often differentiates between affordances that are perceptible and affordances that are not perceptible, or hidden (Gaver, 1991). If a person notices that s/he can act in the environment in a certain way, this is called a perceptible affordance. A hidden affordance refers to an action possibility that a person fails to notice or does not understand (for example, because of poor design). Our model focuses on *user perceptions*, in which, obviously, perceptible affordances play a role. Hidden affordances, therefore, fall outside the scope of our model.

Like the other encode factors in the I-PEFiC model, affordances have a positive and negative dimension (aid versus obstacle, see Figure 3.1). Affordances perceived as offering help (hence, aids) can be used to increase the likelihood that a desired goal can be reached. They indicate that progress is occurring (Peterson, 2005) and evoke intentions to use (cf. technology acceptance theory, Venkatesh et al., 2003). This process is further supported by positive outcome expectancies (i.e., positive valence - goals are supported). The side-effect of this process can be the excitement of positive emotions, such as pleasure and pride (cf. Hoorn and Konijn, 2003). However, affordances can also obstruct goal achievement (hence, obstacles), for example, when the user is in a hurry to finish a document and Clippit pops up with an unhelpful suggestion. Thus, obstacles have the reverse nature of aids. They indicate that the current path of actions chosen may not lead to goal fulfillment (decreased effectiveness) (Peterson, 2005), may prolong goal achievement (decreased efficiency) and/or increase the mental or physical effort required to accomplish a goal. As the technology acceptance theory (Davis, 1989; Venkatesh et al., 2003) argues, use intentions are likely to be influenced by efficiency and effectiveness considerations. Hence, obstacles normally invoke negative valence and result in intentions not to use. This process is accompanied by negative emotions (such as anger and disappointment) as a byproduct (cf. Hoorn and Konijn, 2003).

Most often, positive dimensions (e.g., beauty, aids) will result in positive effects (involvement, intentions to use, satisfaction), and negative dimensions (e.g., ugliness, obstacles) in negative effects (distance, intentions not to use, dissatisfaction). Further, the model allows for *positive effects of negative dimensions* as well as *negative effects of positive dimensions*. Thus, the model also explains why embodied agents that are ugly (negative dimension) can still be involving (positive effect). Or why, in game contexts for example, obstacles might be needed for a positive experience such as challenge, and evoke intentions to use the agent.

3.2.3 Dependencies between the engagement and interaction process

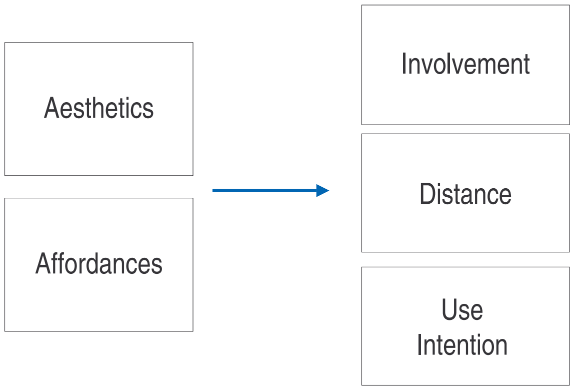
Typically, the appearance aspect of the embodied agent system evokes engagement (involvement and/or distance tendencies), whereas the interactive component (i.e., affordances) yields intentions to use or not. Studies on engagement processes and interaction processes typically have different roots (i.e. communication versus computer sciences, respectively). Because both can occur at the same time and in parallel, the question arises whether these are separate processes or whether interdependencies, such as those modeled in I-PEFiC, feature in human-agent interaction.

First, affordances may influence the engagement process. When an embodied agent is ugly but helpful for task completion (e.g., is efficient), users can still feel involved. I-PEFiC explains this as follows: An affordance in an embodied agent system may increase intentions to use. However, because valence is central to both processes (see Figure 1) and mediates this inclination, part of the influence of affordance may be redirected to involvement towards the embodied agent. Consequently, affordances of the embodied agent system not only affects use intentions but also engagement with the embodied agent. In this case, we speak of 'affective affordances'. Second, appearance aspects, such as aesthetics, may influence the interaction process. When an embodied agent obstructs task completion yet is nicely designed, individuals may still want to use it. In terms of I-PEFiC, aesthetics in an embodied agent may increase involvement, but because this inclination is mediated through valence, part of aesthetics' influence may be redirected towards intentions to use the agent. In other words, an aesthetically beautiful agent may increase not only involvement with the embodied agent but intentions to use as well (cf. 'What is beautiful is usable', Tractinsky et al., 2000). By studying the interdependencies between the processes we may find answers to the following questions: What happens when a user is confronted with an efficient but ugly embodied agent, or, what happens when a user is confronted with a beautiful embodied agent that slows work progress?

If the interaction and engagement processes are unrelated, designers do not have to worry about 'dressing up a system nicely' to motivate individuals to use it. If the system works well, appearance aspects would not affect intentions to use. On the other hand, if the Interaction and engagement processes depend on one another, appearance aspects would affect use intentions and vice versa. Designers could camouflage a deficient agent system by making it look attractive. Conversely, a functional agent system that appears ugly may be misjudged as unusable, which would evoke Intentions not to Use it. Similarly, an agent system that is not usable might lead to felt distance to the agent (cf. Microsoft's Clippit), and a usable agent system may increase involvement despite the agent's outer appearance.

Thus, the challenge is whether we should consider the interactive aspects of bonding with embodied agents as a separate process that is independent of appearance aspects of the agent, or whether it is integral to the embodied agent system experience. We focused on affordances as generators of the interaction process, and

Figure 3.2: Interdependencies between the engagement and interaction processes. Affordances may influence involvement and distance, and aesthetics may influence use intentions (see text).



on aesthetics as producers of the engagement process. Earlier findings (Van Vugt et al., 2005) have shown that (perceived) aesthetics was a stronger predictor of user engagement in human-agent interaction than (perceived) epistemics (realism) of the agent. Remarkably, embodied agent research belittled the importance of aesthetics. Further, other human-computer interaction (HCI) literature indicates that aesthetics has major influence on experiences when users interact with computerized products in general (e.g., Tractinsky, 1997). A confrontation with a (new) system is often a visual one, and during system interaction, visual information constantly is present and immediately evokes aesthetic judgments.

A simple graphical representation of the study can be seen in Figure 3.2.

3.2.4 Hypotheses

Hypotheses derived from the above theorizing were as follows:

H1. If an engagement process is triggered, there is a significant main effect of aesthetics on engagement. If an interaction process is triggered, there is a significant main effect of affordances on use intentions. In both cases, both relevance and valence may serve as a mediator between encode and response factors.

H2. If the engagement process is *independent* of the interaction process then there are no cross-over effects between the two processes. More specifically, there is no main effect of affordances on engagement, no main effect of aesthetics on use intentions, and no interaction effect of affordances and aesthetics on engagement and/or use intentions.

H3. If the engagement process and the interaction process are relatively *dependent* on one another, then, apart from the main effects of H1, cross-over effect(s) will be found. Thus, a main effect of affordances on engagement, a main effect of aesthetics on use intentions and/or a significant statistical interaction between aesthetics and affordances on engagement and/or use intentions should take place. Herein, relevance and valence may serve as a mediator between encode and response factors.

H4. If the engagement and interaction processes work together, this is reflected in end-user satisfaction. Satisfaction, then, depends on both engagement as well as use intentions.

3.3 Method

To verify whether engagement and interaction processes affect each other, we conducted an experiment with the Sims2 game environment. Sims2 offers possibilities to manipulate the interaction through affordances that help achieve a user goal or affordances that hinder achieving that goal. Moreover, Sims2 offers possibilities to adapt the outer appearance of the agents that feature in the game.

3.3.1 Design

Participants (N = 120) were randomly assigned to one of the experimental conditions of a 2 (designed aesthetics: beautiful agent versus ugly agent) x 2 (designed affordances: aid versus obstacle) between subjects design[2].

We implemented the affordance conditions in terms of time-efficiency. That is, in the 'aid' condition, subjects could quickly complete the tasks using the embodied agent. In the 'obstacle' condition, the tasks took at least twice as long. The tasks were designed such that task-completion time was highly dependent on use of the particular embodied agent. For the means and standard deviations of manipulated total task-completion time, see Table 3.1.

Table 3.1: Total task-completion times over the 5 experimental tasks (N = 80), in the two affordance conditions.

	M (sec)	*SD* (sec)	*N*
Aid	302	52	40
Obstacle	675	15	40

We systematically combined beautiful and ugly agents (factor designed aesthetics) with designed affordances that aided achieving a goal (accomplish a task efficiently) versus designed affordances that kept the user from doing so. Four main types of agents were designed: 1) beautiful agents that accomplished tasks efficiently, 2) beautiful agents that accomplished tasks inefficiently, 3) ugly agents

[2]Initially, the design included a half aid - half obstacle condition to study boundary effects. However, the results regarding this condition obscured rather than clarified results. Therefore, we left them out of the present paper.

that accomplished tasks efficiently, and 4) ugly agents that accomplished tasks inefficiently. For dependent variables, we measured involvement with and distance towards the agent (engagement), as well as the intentions (not) to use the agent in a subsequent task.

3.3.2 Participants and Procedure

A sample of 120 university students (69 male, 51 female, mean age 23.97, *SD* = 4.75) participated as volunteers in the study. They were approached on an individual basis and randomly assigned to the experimental conditions. All were experienced computer users, and most of them (86%) had no prior Sims experience.

In a silent room where only one participant and one experimenter were present, the experimenter wrote down personal information, and introduced the Sims game. Then, the participants were informed about the nature of the tasks and were told to perform the tasks as fast as possible to increase the relevance of the task. To keep the participants motivated, the three fastest participants would be awarded with a vacation check worth 100 euros, the Sims2 game worth about 40 euros, and a book check worth 25 euros. They were told to press the stop button of a clock as fast as possible after each task. Then, they were shown a scheme on a second computer and were told that this scheme indicated the average time the tasks were performed by other participants. We made the experimental trials such, that in the aid condition, participants would perform faster than average, and in the obstacle conditions, participants would perform slower than average. Then, the participants were given a practice task to get familiar with the Sims interface and with the clock. If participants had no further questions, the 5 experimental tasks started. For each task, 1) the experimenter gave a short, written, task instruction to the participant, 2) the experimenter started the clock, 3) the participant performed the task and stopped the clock when the task was finished, and 4) the participant wrote down the time in the scheme on the second computer.

After 5 tasks, the participants completed the user-perception questionnaire. In the questionnaire, a final task with the Sims was introduced and they were informed that prices could still be won. Then, the user perception questionnaire asked about the participants' preference to use the same Sim again or another Sim (use intentions). Then, one final (dummy) task was performed and the participants were thanked for participating in the experiment. The prices were given to three students by raffle.

3.3.3 Materials

The Sims2 game, developed by Maxis[3], is one of the best-selling games in computer history. It is a dynamic 3D interactive environment in which human-like agents (the 'Sims') live in a neighborhood, have houses, go to work, sleep, have families and friends, etc. The game options were saved so that all participants

[3] http://www.maxis.com

Figure 3.3: Up, left: Affordances are depicted in the form of menu options next to a Sim. Up, right: Bella is performing the task 'clean up the kitchen sink'. Below: Sims' personality and skills manipulated to create affordances.

started in the same environment. An advantage of the game was that people could easily operate it, with some initial instructions and explanation.

When Sims' 'free will' is turned off, they cannot act on their own. Users could interact with Sims to let them perform certain actions. For example, a Sim could be ordered to read a book or take a shower. The upper left picture in Figure 3.3 shows how users interacted with the agent (let a Sim perform an action) by choosing between menu options (the affordances).

We implemented 5 experimental tasks in which the participants had to let a Sim 1) play the piano (see upper left picture in Figure 3.3, left picture), 2) clean up the kitchen sink (see upper right picture in Figure 3.3, right picture), 3) learn from a book, 4) repair the washbasin, 5) paint on an easel. The order of the tasks was randomized for each participant. Paper instructions explained the tasks one by one, and explained the stop-condition. An example of such an instruction is: 'Bella is standing in front of the kitchen sink. Clean up the kitchen sink. The task is finished when Bella stops cleaning up and stands still in front of the kitchen sink.'

In Sims2, characteristics such as personality traits and skills (see pictures at the bottom in Figure 3.3) influence whether a Sim wants to behave or not. We implemented aids and obstacles by adjusting these characteristics, which in turn affected task-completion time. For example, a sloppy personality and a low cleaning skill may double the time it takes to complete a cleaning task, compared to a neat personality and a high cleaning skill. Similarly, a Sim unskilled in mechanics takes much longer to repair a washbasin than a skilled Sim.

Figure 3.4: Bella (left) and Berta (right) were perceived as beautiful and ugly agents.

The aesthetics of the Sim was manipulated by applying standards of beauty universal in people (Johnston and Oliver-Rodriguez, 1997) to agents, as far as the Sims2 environment allowed. For example, the 'beautiful' agent had a waist-to-hip ratio in the ideal range, an average face shape and symmetry, whereas the 'ugly' agent deviated from these standards. Note that a agent that is visually beautiful differs from a beautiful *design*. An ugly agent can still be designed very well. Therefore, we conducted a pilot study to test whether several female Sims were perceived as beautiful or ugly. The girls presented in Figure 3.4 were the most beautiful (left) and ugly (right) of the set and differed significantly in perceived aesthetics (t(22) = 4.02, $p < .001$, $N = 24$).

3.3.4 Measurements

All measurements were taken by means of a paper-and-pencil questionnaire containing Likert-type scales. Each item was followed by a 6 point rating scale, ranging from 1 (do not agree at all), 2 (do not agree), 3 (barely agree), 4 (agree a little), 5 (agree), to 6 (fully agree). Items were presented in mixed order.

Existing scales were used when possible in the construction of the structured user-perception questionnaire. When necessary, items were translated, adjusted to the purpose of the investigation and the specific material and the language use of the target group of participants (university students). The questionnaire consisted of 48 items.

Reliability analyses ($N = 120$) were performed on each set of items concerning separate scales. Selection criteria were 1) an optimal contribution to Cronbach's alpha by showing little or no increase in the alpha level when the item was deleted, 2) a minimal inter-item correlation of .30, 3) an inter-item total correlation within a scale is bigger than the correlation of each item with another scale (discriminant validity), and 4) a minimum of 2 items per scale. Items that failed on one or more of these criteria were not included in the measurement scales used in subsequent analyses.

We checked for the aesthetics manipulations (designed aesthetics) by means of a *perceived aesthetics* scale. This scale was measured by means of 4 items, based on scales used by Konijn and Hoorn (2005). In order to avoid directing the participant in an affirmative direction (see Dillman, 2000), two items were indicative (e.g., 'Bella looks nice') and two items were contra-indicative (e.g., 'Bella has an ugly appearance'). The scale was reliable according to a Cronbach's alpha of .89.

Further, we checked for the affordance manipulations (designed affordances) by means of a *perceived affordance* scale. Because it was the first time we measured perceived affordance as aids versus obstacles, we used 8 items. Compared to the original scale, one item had to be discarded because of the discriminant validity criterion. The remaining scale consisted of four indicative items (e.g., 'Bella is competent') and three contra-indicative items (e.g., 'Bella is clumsy') and was reliable (Cronbach's alpha = .88).

The measurement of *engagement* was based on scales used by Konijn and Hoorn (2005), distinguishing between *involvement* and *distance*. One item originally designed for the scale did not fit well based on psychometric analyses and was discarded. The remaining involvement scale consisted of three items (e.g., 'Bella appeals to me') and was reliable according to a Cronbach's alpha of .81. The distance scale was measured using four items (e.g., 'Bella leaves me with cold feelings') and was weakly reliable according to a Cronbach's alpha of .59.

The *use intention* measuring scale was based on behavioral intention scales used in technology acceptance literature (e.g., Davis, 1989; Venkatesh et al., 2003)). Factor analyses showed that use intention was a bipolar scale, and did not consist of two unipolar scales as most other I-PEFiC factors. Because of the discriminant validity criterion, one item had to be discarded. The remaining scale existed of four indicative items (e.g., 'I want to use Bella again in a similar task') and three contra-indicative items (e.g., 'For this task, I would have preferred to use another Sim') and was reliable (Cronbach's alpha = .97).

The measuring scale *satisfaction* was based on the scale 'appreciation' used by Konijn and Hoorn (2005), and the ISO 9241-11 standard of satisfaction[4]. Factor analyses showed that satisfaction consisted of two unipolar scales, a positive satisfaction scale and a dissatisfaction scale (just like engagement consists of the scale involvement and distance which are not the ends of one continuum, in Konijn & Hoorn, 2005). Compared to the original positive satisfaction scale, one item had to be discarded because of the discriminant validity criterion. The remaining scale

[4] e.g., http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/ISO9241part11.pdf

existed of four items (e.g., 'Bella is ok') and was reliable (Cronbach's alpha = .81). Compared to the original dissatisfaction scale, three items had to be discarded because of the discriminant validity criterion. The remaining dissatisfaction scale existed of two items ('I am dissatisfied with Bella' and 'Bella makes me sad') and was reliable (Cronbach's alpha = .68).

As in Van Vugt et al. (2005), we measured several additional I-PEFIC variables for their potential influence[5]: *perceived epistemics* (5 items, Cronbach's alpha = .74, after deletion of one item based on psychometric analyses), *perceived ethics* (3 items, Cronbach's alpha = .70, after deletion of one item based on psychometric analyses), *perceived relevance* (4 items, Cronbach's alpha = .79), *perceived valence* (4 items, Cronbach's alpha = .76), and *perceived similarity* (4 items, Cronbach's alpha = .71). The appendix shows the items of the user perception questionnaire and the reliabilities of each scale.

Finally, additional questions asked for some personal information about the participants such as gender, age, computer experience, game experience and previous Sims experience.

3.4 Results

3.4.1 Manipulation check

We assessed the effectiveness of the aesthetics manipulations (designed aesthetics, beautiful versus ugly agent) and the affordance manipulations (designed affordances, aid versus obstacle), by performing a MANOVA with Perceived Aesthetics and perceived affordances as dependents. The tests of between-subject effects revealed that Bella (M = 4.74, SD = .70) was perceived as more beautiful than Berta (M = 2.98, SD = .90; F(1, 116) = 149.83, $p < .001$, partial η^2 = .56). Second, there was a significant difference in perceived affordances between the aid (M = 4.15, SD = .77) and the obstacle (M = 3.42, SD = .80) condition in the right direction ($F(2, 116)$ = 9.83, $p < .001$, partial η^2 = .15). These main effects thus support our manipulation aims.

3.4.2 Preliminary analyses

Participants in various conditions had similar scores on the I-PEFiC factors perceived ethics and perceived similarity. Most participants (83%) scored 4.5 or higher on the Ethics scale, thus they regarded all four agent types as ethically 'good'. Most participants (84%) scored 3 or lower on the Similarity scale, indicating that they perceived themselves as dissimilar to the Sims agents. Perceived ethics and perceived similarity can therefore be regarded constant variables that cannot explain differences between conditions. Therefore, they will not be used in subsequent analyses. The dissatisfaction scale was both severely skewed and peaked, and was

[5]Measures irrelevant for the present study are not mentioned here.

therefore left out of further analyses. The other I-PEFiC variables satisfied the norms of normal distributions.

Age and previous experience with the Sims game co-varied with the designed affordances (age: χ^2 = 12.17, $p < .007$; Sims experience: χ^2 = 29.78, $p < .019$). None of the other personal characteristics had a significant effect on involvement, distance, or use intentions (according to MANOVA). Therefore, age and Sims experience were included as covariates in hypotheses testing. However, they turned out to not significantly affect the dependent variables ($p > .05$).

3.4.3 Hypotheses testing

To test H1 on the main effects of designed affordances on use intentions and of designed aesthetics on engagement (i.e. involvement and distance), a MANOVA was conducted with designed affordances (aids versus obstacles) and designed aesthetics (beautiful versus ugly) as the between-subject factors. The dependent variables were use intentions, involvement, and distance.

The multivariate test showed significant main effects of both designed affordances and designed aesthetics (designed affordances: Wilks' lambda = .90, F(3, 74) = 2.71, $p < .05$, partial η^2 = .10; designed aesthetics: Wilks' lambda = .79, F(3, 74) = 6.43, $p < .001$, partial η^2 = .21). Furthermore, the interaction was significant (Wilks' lambda = .87, F(3, 74) = 3.81, $p < .013$, partial η^2 = .13).

Univariate F-tests confirmed the obtained multivariate results. More specifically, we found a main effect of designed affordances on use intentions (F(1, 76) = 7.70, $p < .007$, partial η^2 = .09), as hypothesized in H1. Thus, an interaction process was triggered. Participants were more keen to use helpful agents, which aid task completion (M = 4.1; SD = 1.1) than obstructing agents, which are obstacles for task completion (M = 3.4; SD = 1.1). Further, designed aesthetics affected user involvement with the agent (F(1, 76) = 17.89, $p < .001$, partial η^2 = .19). Thus, an engagement process was also triggered. Participants felt more involved with the beautiful agent (M = 3.4; SD = .85) than with the ugly agent (M = 2.6; SD = .82). However, people felt as distant to the beautiful agent (M = 3.1; SD = .70) as to the ugly agent (M = 3.3; SD = .74) ($p = .36$).

These results confirmed our expectations regarding the simultaneous existence of an engagement process and an interaction process. However, in this study, our main goal was to test whether the engagement process was either independent (H2) or dependent (H3) of the interaction process. Therefore, we looked into cross-over effects between the two processes. The same MANOVA could be used for this purpose. First, we investigated whether aesthetics affected the interaction process, by investigating effects of designed aesthetics (beautiful versus ugly) on *use intentions*. We found no main effect of designed aesthetics on use intentions ($p = .55$). This suggests that participants were not keen to use a beautiful agent more than an ugly agent, or vice versa. Also, we did not find a significant interaction effect of designed aesthetics and designed affordances on use intentions ($p = .15$). However, the means indicate an interesting trend as shown in Figure 3.5. When an agent

Table 3.2: Means and Standard Deviations of involvement (I), distance (D) and use intentions (U) in the conditions.

	Beautiful	Ugly	Total
Aid			
I	3.3 (1.0)	2.9 (.88)	3.1 (.95)
D	3.2 (.77)	3.0 (.68)	3.1 (.73)
U	4.3 (.89)	3.8 (1.17)	4.1 (1.1)
Obstacle			
I	3.4 (.68)	2.3 (.62)	2.9 (.86)
D	3.0 (.62)	3.5 (.74)	3.3 (.72)
U	3.3 (1.1)	3.5 (1.1)	3.4 (1.1)
Total			
I	3.4 (.85)	2.6 (.82)	3.0 (.92)
D	3.1 (.70)	3.3 (.74)	3.2 (.72)
U	3.8 (1.1)	3.7 (1.1)	3.8 (1.1)

is obstructing, aesthetics does not matter, as participants are not willing to use the agent anyway. However, when an agent is helpful, aesthetics does seem to matter. Then, it seems that participants are more willing to use a beautiful than an ugly agent. This means that a helpful *and* beautiful agent is most ideal for individuals to use.

Figure 3.5: The effect of designed aesthetics and designed affordances on use intentions.

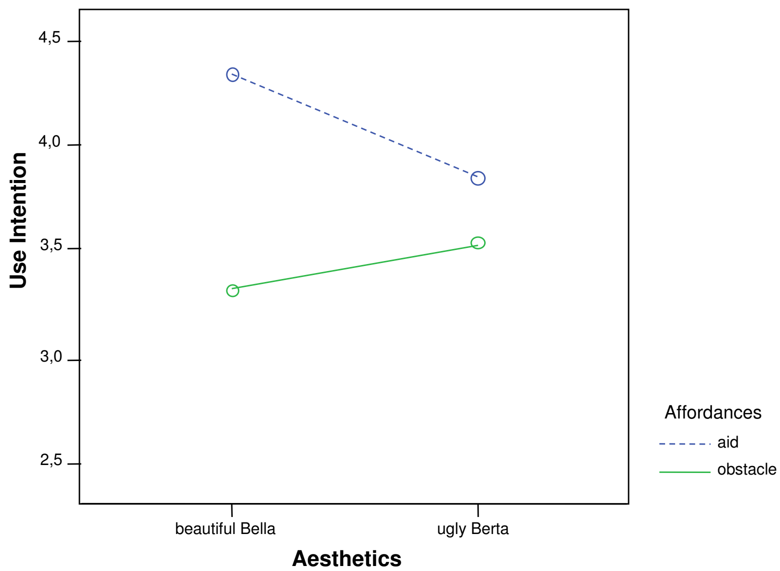
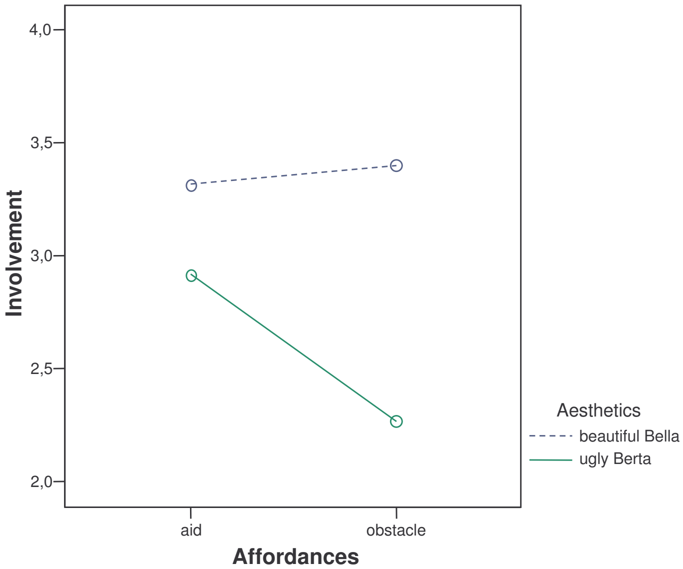


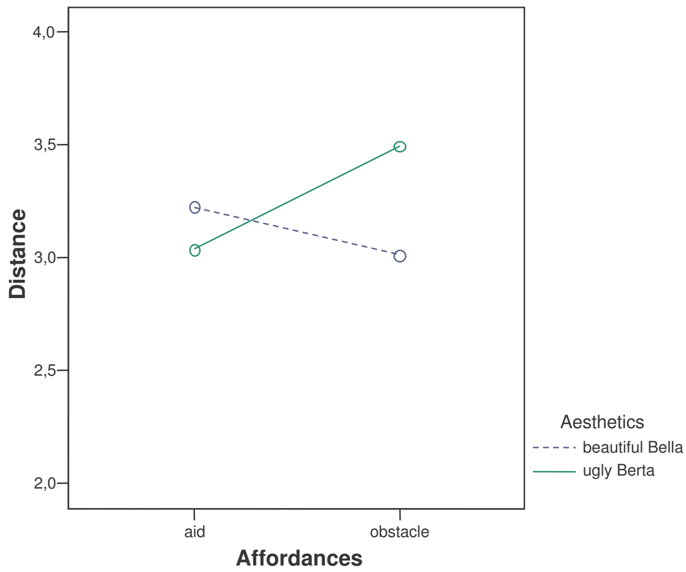
Figure 3.6: The effect of designed aesthetics and designed affordances on involvement.



Second, we investigated whether affordances affected the engagement process by analyzing effects of designed affordances on user involvement and distance. We found no main effect of designed affordances on involvement ($p < .13$), nor a main effect of designed affordances on distance ($p < .44$). However, significant interactions were found on both user involvement ($F(1, 76) = 4.09$, $p < .05$, partial $\eta^2 = .051$) and user distance ($F(1, 76) = 4.47$, $p < .04$, partial $\eta^2 = .056$). Participants were least involved with ($M = 2.3$; $SD = .62$) and most distant to ($M = 3.5$; $SD = .74$) the ugly agent that obstructed task completion (see Table 3.2 and Figures 3.6 and 3.7). Thus, when an agent is beautiful, it does not matter whether the agent helps or obstructs task completion; users are equally engaged with the agent. However, when an agent is ugly, it is important that the agent aids task completion which increases engagement with the agent, despite its ugliness.

To test H4, we used a regression analysis (multiple, method Enter) to predict users' satisfaction with the agent from the continuous variables use intention, involvement, and distance ($R^2 = .34$). The standardized regression coefficients revealed that all three predictors had comparable and significant contributions to satisfaction (use intention: *standardized beta* = .22, $t = 2.70$, $p < .008$, partial r = .24, semi-partial r = .21; involvement: *standardized beta* = .26, $t = 2.84$, $p < .005$, partial r = ..26, semi-partial r = .22; distance: *standardized beta* = -.29, $t = -3.05$, $p < .003$, partial r = -.27, semi-partial r = -.23) These results indicate that end-user satisfaction is a complex construct that depends on both the engagement and the interaction processes.

Figure 3.7: The effect of designed aesthetics and designed affordances on distance.



In sum, the cross-over effects indicate that the two processes are dependent on each other in predicting the agent's effect on its users. Thus, H2, predicting independent processes, was rejected and H3, predicting dependency between engagement and interaction processes, was supported. Likewise, H4 was supported, stating that satisfaction depends on both processes.

3.4.4 Additional analyses

The eye of the beholder. Because the I-PEFiC model predicts that a number of user *perceptions* in the encoding stage are central to user engagement and use intentions, we tested specific user perceptions in predicting user engagement and use intentions. Specifically, we tested whether perceived affordance, perceived aesthetics, and perceived epistemics provided better explanations than designed aesthetics and designed affordances did in the previous analyses[6]. A MANOVA with designed aesthetics and designed affordances as independents, perceived aesthetics, perceived affordance, and perceived epistemics as covariates, and user involvement, user distance and use intentions as dependents, showed significant main effects of factor designed aesthetics (Wilks' lambda = .87, $F(3, 70) = 11.30$, $p <$ .018, partial $\eta^2 = .13$) and all covariates: perceived affordance (Wilks' lambda = .89, $F(3, 70) = 2.98$, $p < .037$, partial $\eta^2 = .11$), perceived aesthetics (Wilks'

[6]The I-PEFiC variable perceived ethics was discarded from analyses because participants rated all agents as equally good.

lambda = .67, $F(3, 70) = 11.30$, $p < .001$, partial $\eta^2 = .33$), and perceived epistemics (Wilks' lambda = .86, $F(3, 70) = 3.72$, $p < .015$, partial $\eta^2 = .14$). Thus, as expected, a variety of user perceptions were important for explaining user engagement and use intentions. Designed affordances and interactions between the factors did not significantly affect the dependents (ps $> .20$). Thus, when users' perceptions of the Sims agents were taken into account, the effect of designed affordances initially found seemed to be overruled. Note, however, that the effects of perceived affordances became significant.

Univariate F-tests confirmed the multivariate results obtained, and indicated the effects of user perceptions specified for the various dependents. Consistent with the multivariate result, the effects of perceived affordance were significant on use intentions ($F(1, 72) = 4.99$, $p < .03$, partial $\eta^2 = .07$), involvement ($F(1, 72) = 3.67$, $p < .06$, partial $\eta^2 = .05$), and distance ($F(1, 72) = 4.05$, $p < .05$, partial $\eta^2 = .05$). The more participants perceived that the agent aided task completion, the more involved and the less distant they felt to the embodied agent. This further supported our hypothesis (H3) that the interaction and engagement processes are dependent on each other.

In addition, we found that perceived aesthetics significantly affected involvement ($F(1, 72) = 18.27$, $p < .001$, partial $\eta^2 = .20$) as well as distance ($F(1, 72) = 24.07$, $p < .001$, partial $\eta^2 = .25$), but not use intentions ($p = .67$). Similarly, perceived epistemics affected involvement ($F(1, 72) = 3.50$, $p < .07$, partial $\eta^2 = .05$) and distance ($F(1, 72) = 10.20$, $p < .002$, partial $\eta^2 = .12$), but not use intentions ($p = .41$). The more beautiful and realistic the participants perceived the agent to be, the more involved and the less distant they felt towards the agent. Thus, perceived aesthetics and perceived epistemics did also play a role in establishing user engagement towards embodied agents, but they did not significantly contribute to use intentions, that is, the interaction process.

In sum, the results of this additional analysis showed that several user perceptions are important in explaining user engagement, namely perceived affordance, perceived aesthetics, and perceived epistemics. We performed a regression analysis to investigate the relative importance of each of these factors. The standardized regression coefficients revealed that perceived aesthetics was the best predictor of user involvement (*standardized beta* = .59), followed by perceived epistemics (*standardized beta* = .30), and perceived affordances (*standardized beta* = .13). Designed aesthetics and designed affordances did not contribute significantly to user involvement. Further, perceived aesthetics was also the best predictor of user distance (*standardized beta* = -.77), followed by designed aesthetics (*standardized beta* = -.43), perceived epistemics (*standardized beta* = .30), and perceived affordances (*standardized beta* = .20). Designed affordances did not contribute significantly to user distance. According to these analyses, a beautiful agent appearance seems crucial to establishing user engagement with an agent. In conclusion, perceptions of the agent's beauty and realistic appearance increase engagement with the agent, but are not that important for use intentions. The agent's perceived affordances also contribute to engagement with the agent and are decisive for use intentions. In assessing end-user satisfaction, however, both the engagement process and interaction process contribute.

Then, from the I-PEFiC model, *in*direct effects of perceived aesthetics and perceived affordances on use intentions and engagement may also be expected. Therefore, we tested the mediating roles of both perceived relevance and perceived valence. We found no effect of perceived aesthetics on use intentions, but we assume this effect is mediated by perceived relevance and perceived valence. To test mediation, the Sobel method is reliable (see Preacher and Hayes, 2004). Contrary to our expectations, tests suggested no mediation effect of perceived relevance (Sobel $z = .54$, $p = .59$) or perceived valence (Sobel $z = -.80$, $p = .42$) between perceived aesthetics and use intentions. Thus, the assumption that perceived aesthetics affect use intentions indirectly, that is, mediated by perceived relevance and perceived valence, cannot be supported.

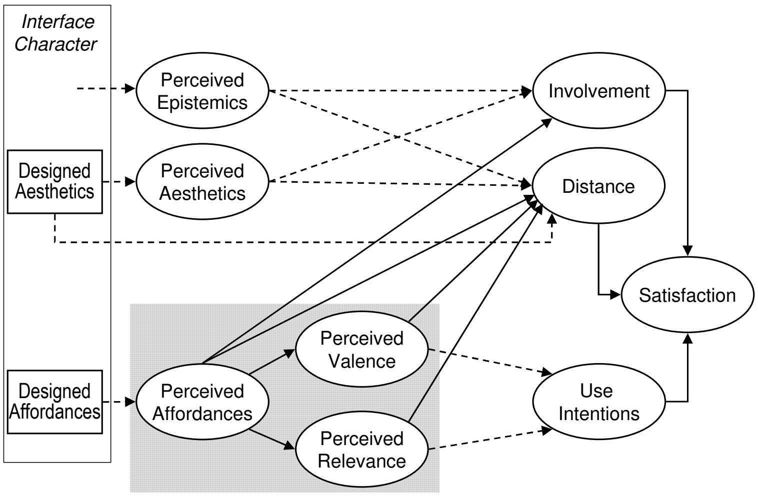
To test a possible mediation effect in the interaction process, we tested mediation by perceived relevance and perceived valence in between perceived affordances and use intentions. As expected, mediation tests revealed that this process is mediated by both perceived relevance (Sobel $z = 3.09$, $p < .002$) and perceived valence (Sobel $z = 3.95$, $p < .001$). This shows that the factors relevance and valence, which are central in the engagement process (Konijn and Hoorn, 2005), are also important in the interaction process.

Last, we tested a possible indirect effect of perceived affordances on engagement. Mediation tests revealed that perceived relevance did not mediate the effect of perceived affordances on involvement (Sobel $z = 1.18$, $p = .24$), nor did perceived valence mediate the effect of perceived affordances on involvement (Sobel $z = 1.09$, $p = .27$). However, the effect of perceived affordances on user distance was significantly mediated by both perceived relevance (Sobel $z = -2.38$, $p < .02$), and perceived valence (Sobel $z = -3.22$, $p < .01$). Thus, perceived affordances affect user distance towards the agent *in*directly, through perceived relevance and valence.

3.5 Conclusion and discussion

Monodisciplinary convention has it that character engagement and user interaction are separate processes. They are studied in separate disciplines (communication science and HCI), and on the face of it do not have much in common. We integrated character engagement and user interaction into the I-PEFiC model (Van Vugt et al., 2004) and in the present paper we demonstrated that the engagement and interaction processes interacted (i.e. exchanged information) while users encountered an embodied agent in a computer task. We established methodologically sound conditions with a commercial off-the-shelf product, Sims2, to guarantee higher ecological validity of laboratory experimentation. Because Sims2 was never developed for manipulation reasons, to our knowledge, such utilization of a commercial game environment is unprecedented.

Figure 3.8: I-PEFiC in the Sims2 case. The arrows indicate relations found in the present study. Drawn arrows indicate the information flow from the interaction process to the engagement process, of which the gray box designates the redistribution center.



The results of our work are summarized in Figure 3.8, which is an adaptation of the previous I-PEFiC model based on the Sims2 data. We did not find significant effects of factors in the engagement process on intentions to use the agent. We found that intentions to use (i.e., the outcome of the interaction process) were mainly dependent on efficiency considerations, and not the agent's visual appearance. Thus, the engagement process did not affect the interaction process. This seems to warrant the observation by Mark (2003) that the computer game industry is shifting from graphical design towards AI gameplay. Probably, a basic level of aesthetics has been established by the graphic designers and the resulting user engagement can probably only be improved by smoothing the affordances as created by AI designers.

Sustaining this suggestion, we found significant effects of factors in the interaction process on engagement. If an agent offered help with task-completion, participants felt more engaged with the agent than when it obstructed task-completion. Hence, *affective affordances* exist. The present study demonstrated that engaging with an embodied agent is connected to the affordances, that is, the agent's action possibilities. This suggests dependency of the engagement on the interaction process (cf. *emotional design*, Norman, 2004). In addition, conforming to our theory and in line with previous research (Van Vugt et al., 2005), we found that a beautiful and realistic appearance increases user engagement. Thus, affective bonds between humans and embodied agents are established by both the agent's visual appearance and the affordances it offers for task execution.

Perceived affordances, thus, was the most important center of redistributing information to the other process. Further, the original I-PEFiC model foresaw that perceived valence and relevance would play mediating roles. Indeed, the effect of perceived affordances on both use intentions and distance was mediated by relevance and valence perceptions. Hence, relevance and valence are also centers of redistributing information (Figure 3.8, gray box).

Further, we found that the parallel and simultaneous engagement and interaction processes both contributed to end-user satisfaction. Satisfied users were those that were engaged with the agent *and* considered the agent helpful. Thus, satisfaction is a subjective sum of several user experiences with the system (c.f. Lindgaard and Dudek, 2003). Because end-user satisfaction is seen as an important goal in user-system interaction and design, our results show that it is important to enhance both engagement and use intentions.

That the engagement process did not influence use intentions has as a consequence that whatever users think emotionally about office applications such as Microsoft's Clippit, their will to use the animated paperclip or not depends solely on the designed and perceived affordances. Making the darn thing look better may do something for the experience, and hence, for user satisfaction, but not for intentions to use it. In other words, covering up a badly designed software product by making it look good may lure a product designer or marketer into thinking that they did a job well done. After all, user satisfaction increased by improving the aesthetic user experience. But what a disappointment if users still leave the application untouched. After all, intentions to use are not fed by mere beauty and thus, obstruction of task execution decreases the intention to use the system in spite of its good looks.

That perceived affordances influenced engagement has another consequence. Whatever users think about the usefulness of a gaming package does affect how they feel about the agents that the game features. Actually, our reasoning should stop here. Statistical rigor should prevent us to go into details about exactly how affordances and aesthetics cooperate in affecting user experience and use intentions. The statistically significant effect we found is that users were most willing to use a agent for task execution when it was helpful and that if it was ugly, it should be helpful to be appreciated at all. However, this is not too exciting common-sense design information. Therefore, we did some eyeball inspection of the means in the Figures 3.5 up to 3.7 and came up with an interpretation that could guide our future research.

3.5.1 Out of the box and into the bin. When pretty embodied agents become helpful or not.

Although the following effects are not significant, the means in Figure 3.5 show a trend that Intentions to Use seemed dependent on efficiency considerations as well as the agent's visual appearance (i.e. designed aesthetics). The speculations offered next may serve as an interpretation of our findings as well as an impetus to explore possible research lines in the future.

Intentions to use the agent for task execution were strongest when the agent was both helpful *and* beautiful. But beauty may be precious; it is also capricious. Users were *least* intending to use a agent when it was unhelpful (understandably) and beautiful (surprisingly)! Yet, this did not harm the user's engagement with the clumsy beauty. Quite the opposite, blundering Bella evoked the highest levels of involvement (Figure 3.6) and the least distance (Figure 3.7). Silly beauties are the nicest but alas, good for nothing.

Silly uglies can count on no mercy. They were deemed least involving and most distancing, yet, seen as somewhat *more* useful for the task than silly beauties. Thus, beauty is a dangerous treasure to cherish. It is a catalyst of extreme judgements. When you are beautiful, you are expected to perform better than others do. If not, you are really a lousy worker but still very nice. Uglies better offer some help to be appreciated. A helpful Berta exerted moderate intentions to use, moderate involvement, and moderate distance. This seems like default attitude that is directed at the most common type of person we encounter; not too beautiful, usually helpful, evoking in us mild feelings of sympathy and reservation.

This becomes all the clearer, considering that the helpful Bella not only raised the strongest Intentions to Use and the strongest involvement, she also raised the same level of or even more *distance* than her ugly but helpful competitor. A top-performing beauty is suspicious, someone to keep an eye on, because she may be a joy but a strong rival as well (cf. Bridget Gregory in Konijn & Hoorn, 2005). Ugliness is more normal, less attractive but also less threatening.

When offered by an embodied agent, then, affordances can become affect-laden (hence, 'affective affordances') and may provoke user reactions other than intended by the sender or designer. If you want to cover up a bad (here slow) system with a beautiful design, you are in deep trouble. Users will like the design a lot but will deem it even worse usable than when you had an ugly (or normal) design. People do not expect too much if it does not look like much (cf. Kurosu and Kashimura, 1995; Tractinsky et al., 2000). The safest route seems to have a plainly looking design with many aiding affordances, so that people are willing to use it and feel mildly involved. If you really want to capitalize on user experience, beware of overdoing it on the side of aesthetics. It may increase the eagerness to use, it may boost involvement, but when it is really helpful to users, they become intimidated and if it does not help them or perhaps is even frustrating to them, your beautiful rubbish system is dragged into the bin.

3.5.2 Methodological considerations for future research

Finding software that allowed researchers to manipulate agents for suit experimental purposes did not appear to be an easy task. Such an endeavor is confined by constraints such as comparability of agents on all features except the manipulated ones, flexibility of agent's features so the experimenter can mold them according to the studies' purposes, and operating in a task-environment where parameters can be set in multiple ways. We found the Sims2 game environment to be very useful for the purposes of the present study. The Sims2 game allowed for manipulations of the task-environment, the tasks to be accomplished, and manipulating the affordances as well as aesthetics factors, whereas the basic characteristics of each version that we created were similar. The Sims2 game suited our experimental purposes surprisingly well. Of course, we also experienced certain limitations of the game's possibilities, such as the limited tasks, the limited relevance of tasks, the limited comparability with tasks and embodied agents as they exist in regular computerized tasks (e.g., a word processor), and the impossibility of performing tasks *without* an agent. Other games' suitability for experimental research on human-agent interactions should be explored. In addition, future research should ascertain whether trends found in the present study are significant in other task-contexts (e.g., entertainment or learning).

Most participants in our experiment had no prior experience with the Sims game and can thus be regarded novices. Experienced Sims users might have perceived the agents and affordances differently than the novices. For example, they might have learned what affordances belong to particular agents shaping their perceptions accordingly (e.g., the valence they attribute to the agent). Long-term investigations might tell us how user perceptions develop over time, and whether interdependencies between the engagement and interaction processes increase or decrease over time. For example, do experienced users only care for task-efficiency, or are they still influenced by beautifully designed agents?

Finally, it would be interesting to measure not only use intentions but actual use as well. To study actual use, we would ideally create an environment in which users have the possibility to turn the embodied agent on and off. Future research might focus on a variety of task-contexts and applications, such as standard desktop applications and the Internet, to further study the relations between the engagement and interaction processes.

In sum, to fully understand user reactions to embodied agents, we recommend a fine-grained perspective, as several factors with roots in different scientific areas contribute. The I-PEFiC model allowed us to understand relations between user engagement and use intentions, factors that are often studied in isolation. Integrating approaches from communication science (i.e., media entertainment theory) with HCI was beneficial, and the study presented in this paper provides us with a more comprehensive understanding of human-agent interactions.

3.6 Acknowledgments

We are grateful to Gerrit van der Veer and Jan Kleinnijenhuis for their generous support and content-related discussions. We thank Egon van den Broek and Piet van Vugt for fruitful discussions in the early stages of this project. Further, we thank Nathalie Vijlbrief for assisting in the experiment.

Appendix A User perception questionnaire

Table 3.3: The questionnaire items (translated from Dutch) of the user-perception questionnaire. The items printed in italics were removed from their scales after scale analysis. Also, the reliabilities of the scales are shown.

Scale	**Reliability** (Cronbach's alpha)	Items	
Aesthetics	.888	Bella looks nice	Bella is pretty
		Bella is an ugly appearance	Bella is nasty to see
Affordance	.882	Bella is competent	Bella is knowledgeable
		Bella is skillful	Bella is clever
		Bella is clumsy	*Bella comes short*
		Bella is a bungler	Bella messes about with things
Epistemics (Realism)	.736	Bella has a natural look	Bella resembles a real life person
		Bella is just like a real person	*Bella is made with fantasy*
		Bella looks fake	Bella differs from a real life person
Ethics	.651	Bella is good-natured	Bella is reliable
		Bella is malicious	Bella is a mean character
Valence	.757	With Bella I make a chance to winn a price	Bella will allow me to perform the next task quickly
		With Bella it will take long to perform the next task	With Bella I will fail
Relevance	.788	Bella is useful in carrying out the tasks	Bella is worthwhile in carrying out the tasks
		Bella is worthless in carrying out the tasks	Bella is useless in carrying out the tasks
Similarity	.709	Bella and I resemble each other internally	Bella and I have characteristics in common
		Bella is different than I am	Bella differs from who I am
Involvement	.808	Bella appeals to me	Bella gives me a good feeling
		Bella attracts me	*I am well-disposed towards Bella*
Distance	.594	Bella leaves me with cold feelings	I feel negatively towards Bella
		I feel distance between Bella and me	I dislike Bella
Use intentions	.965	I want to use Bella in the following task	I want to continue with Bella in the following task
		I want to perform the following task with the help of Bella	I want to work with Bella in the following task
		I'd rather use another Sim to perform in the following task	I want to ignore Bella in the following task
		I want to get rid of Bella in the following task	*It seems like a bad plan to perform the following task with Bella*
Satisfaction	.810	*I am happy with Bella*	It is nice to get on with Bella
		Bella is fine	Bella is ok
		I am satisfied with Bella	
Disatisfaction	.681	*Bella is a failure*	Bella makes me sad
		I think Bella is ridiculous	*I think Bella is a weak Sim*
		I am dissatisfied with Bella	

Bibliography

Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., and Weisbuch, M. (2001). Intelligents agents who wear your face: User's reactions to the virtual self. In A. de Antonio, R. Aylett, D. B., editor, *Intelligent virtual agents*, volume 2190 of *Lecture Notes in Artificial Intelligence*, pages 86–99. Springer.

Berry, D. C., T., L., Butler, and de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, 63:304–327.

Brave, S., Nass, C., and Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62(2):161–178.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH 94*, pages 413–420. ACM Press.

Cassell, J. and Thrisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5):519–538.

Catrambone, R., Stasko, J., and Xiao, J. (2004). ECA as user interface paradigm. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 239–270, Dordrecht / Boston / London. Kluwer.

Champandard, A. J. (2003). *AI game development. Synthetic creatures with learning and reactive behaviors*. New Riders Games.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–339.

Dehn, D. M. and Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. John Wiley and Sons, New York, 2th edition.

Frijda, N. (1988). The laws of emotion. *American Psychologist*, 43(5):349–58.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 79–84. ACM Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gratch, J. and Marsella, S. (2005). Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, 19(3-4):215–233.

Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors (2006). *Proceedings of Intelligent Virtual Agents 2006, IVA'06, Marina del Rey, CA, USA, August 21-23, 2006*, volume 4133 of *Lecture Notes in Artificial Intelligence*. Springer.

Hayes-Roth, B. (2003). What makes characters seem life-like? In *Life-like characters. Tools, affective functions and applications*, pages 447–462. Springer.

Hoorn, J. F. and Konijn, E. A. (2003). Perceiving and experiencing fictional characters: An integrative account. *Japanese Psychological Research*, 45(4):221–239.

Johnston, V. S. and Oliver-Rodriguez, J. C. (1997). Facial beauty and the late positive component of event-related potentials. *Journal of Sex Research*, 34(2):188–198.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A., Bushman, B. J. (2007). World leaders as movie characters? Perceptions of J.W. Bush, T. Blair, O. bin Laden, and S. Hussein. *Media Psychology*, 9(1):157–177.

Kurosu, M. and Kashimura, K. (1995). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference companion on Human factors in computing systems*, pages 292–293. ACM Press.

Lindgaard, G. and Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15(3):429–452.

Mark, D. (2003). An excellent, well-presented look into ai. Spotlight Reviews. Retrieved Nov. 18, 2005 from http://www.amazon.com/.

McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface 2000*, pages 179–186.

Nass, C., Fogg, B. J., and Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6):669–678.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., and Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239.

Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM Press.

Norman, D. A. (2004). *Emotional design. Why we love (or hate) everyday things.* Basic Books, New York.

Persson, P. (1999). Agneta & frida: A narrative experience of the web? AAAI Fall Symposium on Narrative Intelligence.

Peterson, J. B. (2005). The meaning of meaning. In *Selected Proceedings on Searching for Meaning in the New Millennium*. INPM Press. In press.

Picard, R. W. and Klein, J. (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with Computers*, 14:141–169.

Preacher, K. J. and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4):717–731.

Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge University Press.

Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors (2003). *Intelligent Agents, 4th International Workshop, IVA 2003, Kloster Irsee, Germany, September 15-17, 2003, Proceedings*, volume 2792 of *Lecture Notes in Computer Science*. Springer.

Rosis, F. D., Pelachaud, C., Poggi, I., Carofiglio, V., and Carolis, B. D. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59(1-2):81–118.

Ruttkay, Z., Dormann, C., and Noot, H. (2004). Embodied conversational agents on a common ground. a framework for design and evaluation. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 27–66. Kluwer Academic Publishers.

Ruttkay, Z. and Pelachaud, C. (2004). *From brows to trust: Evaluating embodied conversational agents*. Kluwer Academic Publishers, Dordrecht / Boston / London.

Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 115–122. ACM Press.

Tractinsky, N., Shoval-Katz, A., and Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13(2):127–145.

Van Vugt, H. C., Hoorn, J. F., and Konijn, E. A. (2004). Adapting empathic agents to user experiences. In AAMAS ’04 Workshop on Empathic Agents.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., Eliëns, A., and Keur, I. (2005). Emotional engagement with embodied agents. In Workshop on Emotion and Interaction.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

CHAPTER 4

Facial similarity between embodied agents and users: the differential effects of affordances and gender

Abstract
The aim of the present experimental study was to investigate the effects of facial similarity between user and embodied agent on user responses under different conditions. Sixty-four undergraduates interacted in two tasks with a morphed embodied agent in a virtual environment. In one task the agent was facially similar, and in the other the agent was facially dissimilar to the participant. The agent had *aiding* or *obstructing* affordances. Results showed that the facial similarity manipulation affected participants' responses even though they did not consciously detect it. When paired with aiding affordances, facial similarity increased involvement of all participants and boosted the use intentions of male participants only. However, when facial similarity was paired with obstructing affordances, male participants preferred to use a facially *dis*similar agent. Further, affordances affected the various user responses indirectly, via perceptions of relevance and valence. These results suggest that agents that fail to perform their task and do not meet user goals better not look like the user. We provide several potential explanations for the differential effects of gender, which may guide future research.

This chapter is resubmitted for publication:
Van Vugt, H. C., Bailenson. J. B., Hoorn, J. F., and Konijn, E. A. (submitted). Facial similarity between embodied agents and users: the differential effects of affordances and gender.
A previous version of this chapter is accepted for presentation:
Van Vugt, H. C., Bailenson. J. B., Hoorn, J. F., and Konijn, E. A. (2008). Love my face in virtual space, unless I am male and he is incompetent. To be presented at the 58th annual conference of the International Communication Association (ICA'08), Montreal.

4.1 Introduction

Creating an embodied agent with a three-dimensional face that looks like the face of the user is a relatively new application area in the field of Human-Computer-Interaction (HCI). Due to recent technological advancements, such systems may soon be on the market. Consequently, it is important to understand how users respond to self-similar agents. However, whereas the realism of agents received much attention (Yee et al., 2007), research on facial similarity in embodied agents is relatively new and scarce.

Physical appearance is an accessible and salient dimension along which people judge others (e.g., Sangrador and Yela, 2000), and the face is one of the most important dimensions involved in social interactions (e.g., Sergent and Signoret, 1992). Facial characteristics may affect people ‘in a glance’. For example, people can be instantly and unconsciously affected by facial attractiveness (Langlois et al., 1991; Olson and Marshuetz, 2005), as well as by the *similarity* between their own face and the face of another person. Often, facial similarity ‘attracts’. For example, people tend to marry people who look like themselves. Hinsz (1989) showed participants photographs of actual or randomly generated couples and asked them to rate the similarity between the faces. Actual couples showed higher facial similarities than the ad-hoc couples. People also tend to have pets that look like themselves. Payne and Jaffe (2005) showed that in a significant proportion of human-pet pairs sampled in pet beauty contests, the partners showed much higher facial resemblances than could be expected by random pair formation. Facial similarity may even affect people’s voting behavior during elections. Bailenson et al. (2008) morphed faces of voters with faces of president candidates and found that, although strong partisans were unmoved by a facial similarity manipulation, weak partisans and independents preferred the candidate with whom their own face had been morphed over the candidate morphed with an unfamiliar face. Thus, facial similarity is an important determinant of people’s judgments of and responses to others.

The Computers Are Social Actors paradigm (Reeves and Nass, 1996) states that people respond to computers as they do to real people. Thus, one would expect such effects to occur when people interact with embodied agents that feature in computer environments as well. Embodied agents are computer programs that appear on a computer screen with a life-like visual appearance (e.g., Dehn and Van Mulken, 2000; Cassell et al., 2000; Prendinger and Ishizuka, 2004; Ruttkay and Pelachaud, 2004; Gratch and Marsella, 2005). This could be on the Internet (e.g., ANNA on the IKEA website[1]), in productivity applications (e.g., Clippy in Microsoft applications) and in virtual worlds (e.g., Grand Theft Auto). Research has shown that people indeed are affected by facial similarity when they interact with embodied agents. First, Nass et al. (1998) had participants evaluate a system that displayed an autonomously behaving agent that looked liked them or like someone else. Results showed that when seeing the self-image, participants

[1] http://www.ikea.nl, retrieved January 10th 2008

claimed more responsibility for the evaluation and perceived the evaluation to be more valid and objective. A second study, using photographs as agents in a game context, found that participants trusted facially similar agents more than facially dissimilar others (De Bruine, 2002). A third study found that facial similarity can be used to gain social influence, especially when the embodied agent also mimics the user in behavior (Ratan and Bailenson, 2007). Finally, Li et al. (2007) demonstrated that participants rated the persuasiveness of agents to change their choices higher when the agents' faces were facially similar to their own compared to when the faces were dissimilar. Thus, people are affected by facial similarity when they interact with embodied agents. However, unlike the above studies suggest, people may not always prefer to use facially similar over dissimilar agents.

4.1.1 Similarity and social comparisons

Social comparison theory states that people constantly compare themselves to others (Festinger, 1954; Heider, 1946). The face, gender, ethnicity, personality, and attitude are examples of dimensions on which people may compare themselves to others, including embodied agents (concerning agents, e.g., Dryer, 1999; Nass and Moon, 2000; Nowak and Rauh, 2005; Guadagno et al., 2007; Pratt et al., 2007). One important motive for social comparisons is that people want to make accurate self-evaluations (Festinger, 1954). Thus, when people compare themselves with a person with attributes similar to theirs, they unconsciously 'look in the mirror' and evaluate not only the other person or agent, but also themselves.

There is abundant evidence that, more often than not, 'similarity attracts' (e.g., Byrne, 1971; Chaiken, 1979; Brock, 1965; Cialdini, 2001; Klohnen and Luo, 2003). In such cases, social comparison with similar others may have positive consequences for self-evaluation. Based on the above studies, one would simply expect that people always prefer to use facially similar over dissimilar agents. However, social comparison with similar others may not only result in attraction but also in rejection (Lerner and Agar, 1972). Likewise, people may not always approach similar others, but also avoid them [ibid]. The reason is that there are other, *negative*, characteristics involved in assessing other social actors. Based on the negative characteristics of the similar other, people (unconsciously) activate negative self-associations, and feel threatened in their personal identities (e.g., Lerner and Agar, 1972). One reason might be that people feel that interpersonal similarity on one dimension, such as personality, implies similarity on other dimensions (Heider, 1946), including those perceived as unfavorable. Harshly evaluating a similar but unfavorable other indicates dissociation and reduces the chance for people of being cast in the same negative light (Eidelman and Biernat, 2003). Thus, individuals devaluate a similar unfavorable other as a protection strategy that distances themselves from the other [ibid]. Indeed, a few studies have demonstrated that, under certain conditions, people are less favorable towards similar others than towards dissimilar others. Particularly, when similarity is paired with negative characteristics, such as unattractiveness, mental disturbance, or obnoxious behavior, similarity may have negative rather than positive implications. Taylor and Mettee (1971), for example, created similarity conditions by matching or mismatching personality traits of a participant and another person. When similar others were portrayed as pleasant, they were preferred over dissimilar others. However, when similar others were portrayed as obnoxious, participants preferred dissimilar others. Another study tested the effects of attitude similarity concerning prison reform (Silvia et al., 2005). Republicans and Democrats listed either their own similarities to or differences from prisoners, and then completed a measure of prison reform attitudes. Findings showed that Democrats (a group with positive attitudes toward liberal prison reform) became even more positive when they considered their similarities to prisoners. Republicans (a group with negative attitudes toward liberal prison reform), however, became even more *negative* when they considered their similarities to prisoners. It seems then that in itself facial similarity evokes positive effects but that it can easily be overruled by characteristics that induce negative affect.

Because we are implementing facial similarity as part of a software agent, *affordance evaluations* are central. Affordances are the possibilities for action that the software offers to the user (search, help, etc.) (cf. McGrenere and Ho, 2000; Gibson, 1979). The help function of Microsoft's Clippy can always be called upon irrespective of the user's potential needs and goals. In this manner, the affordances an agent offers can be perceived by the user as either aids or as obstacles, depending on the goal context (Van Vugt et al., 2006). If Clippy's suggestions are correct, Clippy is likely to be perceived as having aiding affordances by users who seek advice from the agent. If Clippy makes suggestions that are beside the point or incorrect, its affordances are likely to be perceived as obstructive. An obstructive affordance typically is perceived as a negative characteristic of an embodied software agent, which might be detrimental to the positive effects of being facially similar to the user. All the literature that reported positive effects of facial similarity used agents that were helpful to, or at least not obstructing, the user's goal. However, users may prefer *dis*similarity over similarity, especially when they interact with agents they perceive as obstructing. Users may even be less willing to interact with obstructing agents that look like them than with obstructing agents that are dissimilar. Therefore, we studied facial similarity and affordances of embodied agents in parallel, and studied their combined effects on user responses.

4.1.2 User responses: involvement, distance, and use intentions

We studied the combined effects of facial similarity and affordances on two types of measures. First, we measured the user's psychological *involvement* with and psychological *distance* towards the agent, which are related to user affect. Involvement refers to psychological approach tendencies (e.g., empathy, sympathy, challenge), and distance refers to psychological avoidance tendencies (e.g., antipathy, irritation, boredom). Traditional views on approach-withdrawal processes (e.g., Russell and Carroll, 1999) consider involvement as an opposing feeling of distance. However, recent research in attitudinal ambivalence has claimed that the affect system should better be conceived of as separate orthogonal positive and negative substrates (e.g., Cacioppo and Berntson, 1994; Diener and Emmons, 1985; Priester and Petty, 2001). We assumed, therefore, that two unipolar constructs of involving and distancing processes toward agents is preferable to the common bipolarity (for a more detailed explanation, see Konijn and Hoorn (2005)). Indeed, earlier research repeatedly found that experiences of involvement occur parallel to experiences of psychological distance (Konijn and Hoorn, 2005; Konijn and Van Vugt, 2008). For example, people may experience both involvement and distance when an evidently good guy shows a dark side (e.g., Batman). Similarly, a user might feel sympathy for an agent and find the agent boring at the same time (e.g., a virtual news-reader). People can feel very attracted to an embodied agent because of its beautiful appearance and at the same time feel very distant from it because that appearance appears to cover spy ware (cf. 'Virtual Katja'[2]). Because researchers are still debating whether positive and negative affect should be treated as one factor or as two separate factors (e.g., Brehm and Miron, 2006), we will treat involvement and distance as separate variables in the present study, and we will test whether this distinction is valid. For simplicity's sake, our hypotheses predict that involvement and distance are a mirror image of one another. Due to their unipolar nature, however, it may also turn out that the effects on involvement are absent whereas the effects on distance are significant, or vice versa.

Next to the involvement and distance measures that are related to user affect, we also measured intentions to use the agent. Intentions have been used in previous work as a proxy for behavior in general and for usage behavior in particular (e.g., Fishbein and Ajzen, 1975; Ajzen, 1991; Davis, 1989; Venkatesh et al., 2003). Affective responses (e.g., involvement, distance) and behavior (e.g., use intentions, actual use) do not always go hand in hand. For example, many features of an intranet system that were actually used, actually also were disliked (Saffer, 2007). Similarly, when users feel involved with an embodied agent 'on a personal level', they may still decide not to use it because it is inefficient to do so (e.g., in the case of misplaced pro-active behavior of Clippy). People may clearly feel psychologically distant to an agent but may still use it, because they cannot perform a task any other way (e.g., buy a difficult to find product via a product presenter agent). Therefore, we treat involvement, distance, and use intentions as separate factors that are not necessarily interconnected.

The quality of affordances is an important predictor of both involvement, distance and use intentions (Van Vugt et al., 2006). That is, an embodied agent with aiding affordances evokes high involvement and low distance, as well as intentions to use the agent. An embodied agent with obstructing affordances evokes low involvement and high distance, as well as intentions *not* to use the agent. The purpose of the current study is to examine in how far facial similarity can boost or counter such general effects. Therefore, we expected that when people are confronted with an embodied agent with aiding affordances, facial similarity will increase involve-

[2]http://www.katja-schuurman.com, Retrieved January 1st 2008

ment with and intentions to use the agent, and decrease the distance felt towards the agent. When people are confronted with an embodied agent with *obstructing* affordances, facial similarity will *decrease* involvement with and intentions to use the agent, and *increase* the distance felt towards the agent. In the second case, facial dissimilarity is preferred to facial similarity. Thus, we will test the following hypotheses:

> **H1** When an embodied agent has aiding affordances, facial *similarity* leads to
>
> (a) higher use intentions, and
>
> (b) higher involvement / lower distance
>
> than facial dissimilarity.
>
> **H2** When an embodied agent has obstructing affordances, facial *dis*similarity leads to
>
> (a) higher use intentions, and
>
> (b) higher involvement / lower distance
>
> than facial similarity.

Thus, we expected that under certain circumstances (H2), use intentions and involvement are higher when the embodied agent does *not* resemble the user than when it does.

4.1.3 Differential effects of user characteristics

How users respond to an (facially similar) embodied agent does not only depend on the affordances an agent provides to the user, but also on a variety of user characteristics, such as gender, age, ethnicity, education, computer experience, and others (e.g., Catrambone et al., 2004; Ruttkay et al., 2004). For example, unskilled Word users might find Clippy helpful when it gives text-editing suggestions, whereas skilled Word users might find Clippy's suggestions disturbing for the tasks they want to perform. Or, a child might like to see the cute dog or the lovely cat that blinks at him or her, whereas adults might not find it entertaining but rather annoying. One variable which has been examined specifically in relation to facial similarity is gender. De Bruine (2004a) found that facial resemblance increases the attractiveness of same-gender faces more than opposite-gender faces. The effect occurs for both men and women who interact with those faces (De Bruine, 2002, 2004b,a). Similarly, Nass et al. (1998), Li et al. (2007), and Ratan and Bailenson (2007) did not report any gender or age differences in their studies on facial similarity effects. However, in terms of interaction with agents in general, there has been work indicating that men and women interact differently with embodied agents on a social level. For example, in two studies, women showed greater conformity with agents than men (Lee, 2003, 2007), and in another study, female participants were affected by mutual gaze behavior of agents, whereas male participants were not (Bailenson et al., 2001). In (Bailenson et al., 2005), men recalled more verbal information from a persuasive message given by an agent than women. Because the literature is unclear in this regard, we did not make predictions concerning differences in gender or other user characteristics. Nevertheless, we planned to control for these user characteristics in our experimental study, and examined the data separately by gender.

Another user characteristic which varies across individuals is *goals and needs*. Theories in psychology describe that, to a considerable extent, human action is goal-driven (e.g., Leontiev, 1978; Gollwitzer and Bargh, 1996). People typically use an affordance because of a goal they want to achieve (e.g., finding information, having fun). Several strains of research are based on the idea that goals are central to human activity, in, for example, the communication sciences (e.g., Blumler and Katz, 1974) and the field of human-computer interaction (e.g., Card et al., 1983; Van Welie and Van der Veer, 2004; Kaptelinin and Nardi, 2006). The user-centered approach to designing computer systems (e.g., Norman, 1988; Preece et al., 2002) also stresses the importance of looking into human factors, such as user goals, in order to create well-designed systems that users care for and use. For example, to assess the impact that information retrieval systems have on users, the classical precision and recall measures are ineffective, and measures of situational *relevance*, which consider the impact of the system on the user, are proposed (Harter, 1992; Hersh, 1994). Theories from psychology (Frijda, 1986, 1988; Lazarus, 1991) state that the *strength* of a human response is guided by the relevance of particular features of a stimulus to the human goals or needs. Similarly, an agent that is perceived as relevant by the user, is likely to evoke stronger user responses (e.g., higher use intentions) than an agent that is perceived as irrelevant (cf. Ajzen, 1991; Venkatesh et al., 2003). Furthermore, the *direction* of human responses (positive versus negative) is a function of *valence* (Frijda, 1986, 1988; Lazarus, 1991). In an embodied agent context, valence is the expectation of whether using the agent will lead to achieving user goals or hindering them (cf. House and Perney, 1974; Van Vugt et al., 2006). Valence is positive when the user's expectations of using the agent will lead to achieving user goals. Valence is negative when the user's expectations on using the agent will lead to leading away from user goals. In general, relevance and expected goal achievement leads to a tendency to use the agent, whereas irrelevance and expected goal evasion leads to the tendency to stay aloof (cf. Ajzen, 1991; Venkatesh et al., 2003; Konijn and Hoorn, 2005; Van Vugt et al., 2006).

Perceptions of relevance may depend on facial similarity. In the social comparison literature, perceived relevance is an important construct to further understand the effects of (dis)similarity to the observer (Kruglanski and Mayseless, 1990; Tesser, 1988). In general, similarity is not uniformly related to relevance (Kruglanski and Mayseless, 1990). In some cases, a similar other is perceived as more relevant than a dissimilar other. For example, when choosing a novel, a person who likes to read mystery is more likely to seek advice from someone who shares his or her literary taste and preferences, a similar other, than from someone who reads science fiction (Goethals and Darley, 1977). In other cases, a dissimilar other is perceived as more relevant than a similar other. In selecting what film to watch, people might turn to an expert in reviewing films on the Internet or in newspapers, a dissimilar other, because the expert might be more reliable (relevant) than a friend [ibid]. Thus, the perception of relevance might also be key to understanding why similarity has positive or negative effects in virtual worlds.

Perceptions of valence may also depend on facial similarity. People may predict the behavior of a similar agent better than that of a dissimilar agent (cf. Miller and Marks, 1982; Mumford, 1983; Ames, 2004). Because people know what to expect from themselves, they may expect different things from a self-similar agent than from a dissimilar agent.

In the present study, we expected that facial similarity would alter the user's perceptions of relevance and valence, which consequently affect user involvement, psychological distance, and use intentions (cf. Ajzen, 1991; Venkatesh et al., 2003). In other words, relevance and valence may serve as mediators between facial similarity and the various user responses (for a detailed explanation on mediation effects we refer to Baron and Kenny, 1986). If such mediation effects occur, we may infer that user goals are important to better understand facial similarity effects in virtual environments. This may consequently guide future research on facial similarity effects in goal or task contexts. We will test the following hypotheses:

> **H3** The effects of similarity on involvement, distance and/or use intentions are mediated through
>
> (a) perceived relevance
>
> (b) perceived valence

Yet, perceptions of relevance and valence also depend on the affordances of the agent. In a previous study (Van Vugt et al., 2006), we found that the factor affordances *in*directly affected the dependent variables (involvement, distance, and use intentions) via the mediation variables perceived relevance and perceived valence. Users perceived the affordances an agent offered, and then checked whether the affordances were relevant for their goals and whether positive or negative outcomes could be expected from using the agent. This consequently affected user responses, such as intentions to use the agent [ibid]. We expected that such mediation effects were also prevalent in the 3D immersive virtual environment that we used in the present study, and we tested the following hypotheses:

> **H4** The effects of affordances on involvement, distance and/or use intentions are mediated through
>
> (a) perceived relevance
>
> (b) perceived valence

In sum, this study investigated both the (combined) effects of designed features (here, affordances and facial similarity) on the user, and the psychological factors relevance and valence underlying such effects. Note that in the present study, we did not manipulate relevance and valence, but rather studied their mediating effects.

4.2 Method

4.2.1 Overview of the study

The present study examines the (interactive) effects of facial similarity and affordances on use intentions and engagement with embodied agents, and the role of user perceptions herein. Most previous work on facial similarity employed morphing techniques to 2D faces, and Bailenson et al. (2008) provides a review. However, the present study employed morphing techniques to create 3D embodied agents that looked like each of the participants. In addition, the study uses an immersive virtual environment, an excellent research tool for studies on applied social psychology (Blascovich et al., 2002).

Each participant performed two tasks in virtual reality. Usually, virtual reality places *multiple* people in a virtual space in which they are able to see representations of each other, using head-mounted displays (Lanier, 2001). For experimental purposes, the participants in our study did not see multiple people in the virtual world, but only *one* agent. In one task, they interacted with a facially similar embodied agent, and in the other they interacted with a dissimilar embodied agent (*Designed similarity*). Half of the participants interacted twice with an embodied agent that provided good advice, whereas the other half of the participants interacted twice with an embodied agent that provided bad advice (*Designed affordance*). After each task, participants completed a user perception questionnaire, in which we measured, most importantly, (1) the dependent variables (involvement, distance, use intentions), (2) perceptions of relevance and perceptions of valence, (3) whether participants indeed perceived the quality of the advice as aiding versus obstructing (*perceived affordances*), and (4) how much participants perceived, or reported, themselves to be similar to the agent they interacted with (*perceived similarity*). This is in line with researchers studying interpersonal attraction in the context of romantic relationships who have distinguished between the effects of actual similarity and perceived similarity (e.g., Levinger and Breedlove, 1966).

4.2.2 Experimental design

We used a 2 (*Designed similarity*: facial similar versus facial dissimilar) x 2 (*Designed affordance*: aid versus obstacle) experimental design to test our hypotheses. We manipulated designed similarity as a within-subject factor, and designed affordance as a between-subject factor. Participants were randomly assigned to the designed affordance conditions. The order of the facial similarity conditions and the pairing of similarity conditions with other between-subjects conditions was varied according to a counterbalanced scheme. In addition, we used two different names and voices for the two agents with which participants interacted, and these were also counterbalanced among the conditions. The results of hypothesis testing did not change when we controlled for the different voices of the embodied agents, and no order effects were found.

4.2.3 Participants

A sample of 64 university students (22 male, 42 female; *M*(age) = 20.5, *SD*(age) = 3.0) participated in the study for course credit. They were randomly assigned to the experimental conditions. The aid condition consisted of 12 males and 20 females, and the obstacle condition consisted of 10 males and 22 females. Most participants (97%) spent more than 6 hours a week behind a computer, and hence can be regarded as experienced computer users.

4.2.4 Stimuli

Designed similarity

Frontal and profile photographs of 64 undergraduate students were taken in the laboratory during the first week of an academic quarter. Over the next 6 weeks, these photographs were converted into virtual busts i.e., digital, three-dimensional busts using 3DmeNow photogrammetric software, see Bailenson et al. (2004) for a thorough description of this process. See also Figure 4.2. The virtual busts were then 'morphed' with each participant's head bust using both geometric and textural algorithms from Vizard 2.53 (see Figure 4.1). Both geometry and textures were morphed at a 35% (participant) - 65% (embodied agent) ratio because previous studies have demonstrated that this ratio is the optimal balance between manipulation effectiveness and non-detection of the morph (concerning 2D faces, Ratan and Bailenson, 2007; Bailenson et al., 2008). An additional contribution of the study is to test whether 35% is enough to detect effects in 3D virtual environments.

The head was morphed with a three-dimensional head bust of an embodied agent of the same gender. In total, we used four different male and four different female heads (taken and adapted from heads used in a previous study of Yee and Bailenson, 2007) to reduce influences of the particular stimuli faces. A given female participant was morphed with one of the four female heads according to a counterbalanced scheme, which equalized frequency of each head's use. After the morphing process was complete, the virtual busts were imported into the three-dimensional virtual reality scene using Vizard 2.53, and attached to generic male and female bodies (see Figures 4.1 and 4.2).

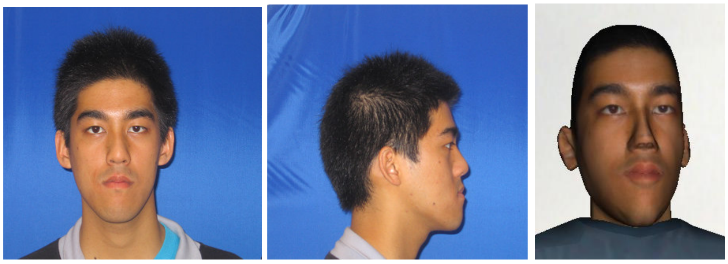


Figure 4.1: A frontal and profile photo of a participant (left and middle) were converted into virtual busts attached to a virtual body (right).

Figure 4.2: The manipulation of facial similarity. An embodied agent head (left) is morphed with a participant head (middle), resulting in a new head (right).

Designed affordance

Many embodied agents aim to provide advice. Therefore, we manipulated affordances in terms of the quality of advice an embodied agent would give to the participants while performing the tasks. Intelligent advice is likely to be perceived as an aid, whereas bad advice is likely to be perceived as an obstacle for task completion. Hence, in the aid condition, the embodied agent would give high quality advice (10 out of 10 questions correct), and in the obstacle condition, the embodied agent would give low quality advice (only 2 out of 10 questions correct). Post-hoc tests confirmed this prediction (see *Manipulation check affordances* in the Results section).

In total, we selected 20 difficult trivia questions with multiple choice answers. Pre-testing demonstrated that very few people knew the answers to these questions; to each of the questions, only 25% guessed the right answer (which equals the chance level). An example question was 'In the U.S., about how many gallons of milk does the average cow give during a year?' with the multiple-choice answers (A) 1600, (B) 1200, (C) 1000, and (D) 1400. Figure 4.3 shows an example of a trivia question, and the embodied agent's advice, which was either correct (see

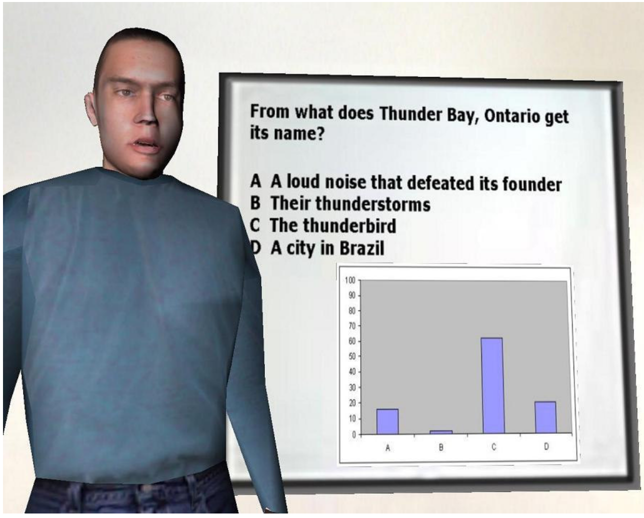


Figure 4.3: An example of an embodied agent giving correct advice (*aid* condition).

Figure) or incorrect. The bars in the graph indicate how certain the agent is of each of the four answers and have different values in the various conditions.

The assignment of heads to conditions

Participants performed two tasks, each with a unique embodied agent as an advisor. Each participant always interacted with two embodied agents of the same affordance condition. In the first task, half of the participants interacted with an agent morphed with their own head, the virtual 'self' (similar condition), and the other half interacted with an embodied agent morphed with another participant's head of the same gender, a virtual 'other' (dissimilar condition). In the second task, this was reversed: participants who had interacted with their virtual self in the first task would interact with a virtual other in the second task, and vice versa. Thus, if a female participant's virtual self was morphed with female agent head 1, the virtual other would be another female participant's head morphed with female agent head 2, 3 or 4. An agent that was morphed with the head of participant A was seen by two participants: participant A (similar condition) and one other participant (dissimilar condition).

The virtual setting

The virtual setting was a white room with the same exact dimensions as the physical room where participants were. The embodied agent was located near the participant, facing the participant and standing next to a blackboard (see Figure 4.3).

It had an automatic blink animation based on human blink behavior and lip movement that matched the volume of its speech.

4.2.5 Apparatus

Participants wore an nVisor SX HMD that featured dual 1280 horizontal by 1024 vertical pixel resolution panels that refreshed at 60 Hz. The display optics presented a visual field subtending approximately 50 degrees horizontally by 38 degrees vertically.

A personal computer with an NVIDIA GeForce FX 6800 graphics Card rendered stereoscopic images for correct perception. These images updated at an average frame rate of 60 Hz. The simulated viewpoint updated continually as a function of the participants' head movements. A three-axis orientation sensing system (Intersense IS250, update rate of 150 Hz) tracked the orientation of the participant's head. The system latency, or delay between a participant's head movement and the resulting concomitant update in the HMD's visual display was 45 ms maximum. Vizard 2.53 software was used to assimilate the rendering and tracking. Participants used a Logitech RumblePad Pro game pad to interact with the virtual environment.

Participants answered post-experiment questionnaires on a standard desktop computer. See Figure 4.4 for equipment setup.

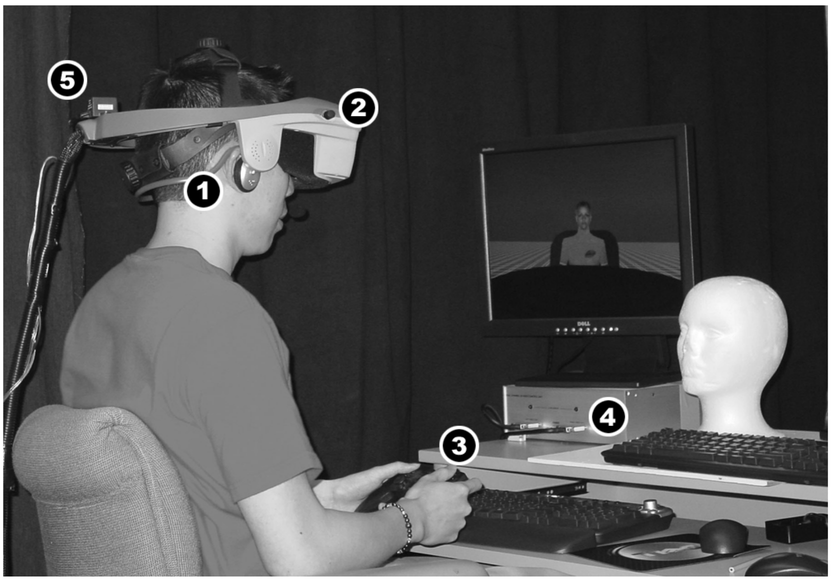


Figure 4.4: A depiction of our immersive virtual environment system. The components are: 1) audio output device, 2) HMD, 3) game pad input device, 4) image generator, and 5) orientation tracking device.

4.2.6 Procedure

Besides the participant, only one experimenter was present in the experimentation room. Participants were told they would perform two simple tasks within a virtual reality environment, and that their task was to complete a questionnaire as best as they could within the virtual environment. Then, they were told the following:

> 'This is unlike most questionnaires where you are on your own. Here, you will have the help of a virtual human. The virtual human is an intelligent agent that is designed to search for information on the Internet. You can choose to use the hints of the virtual human or not. It is completely up to you. In the computer, the correct answers are stored. The computer will tell you after each question whether you answered a question correctly or not. The virtual human does not have access to these correct answers, but reasons him/herself autonomously.'

Then, participants were seated on a stationary chair in the lab, given a game pad and asked to put on the HMD (see Figure 4.4). Participants were then asked by the experimenter to verify that they saw a virtual room. If the participant was ready, the participant would do a practice task within the virtual room to get used to the game pad for answer selection. The experiment began immediately after that. The experiment consisted of two tasks, with 10 questions each.

The embodied agent greeted the participant: 'Hi, my name is Jane [John]. Your task is to respond to ten questions as best as you can. I will be your advisor.' Then, the virtual blackboard displayed with 3 second intervals: 1) a question, 2) the multiple choice answers, 3) a graph with the advice of the embodied agent, and 4) the selection categories A, B, C, and D (see Figure 4.5). At time interval 3, the embodied agent said, 'In the graph you can see what I think the correct answer is. Please select your answer.' The participant would select an answer and press the OK button. Participants got immediate feedback on correctness of the answers (see Figure 4.5). A series of 10 questions were answered in this way. Upon completion of all questions, the embodied agent said: 'Thanks, and maybe I will see you again!' Finally, the participant was asked to remove the HMD and complete the user perception questionnaire at a desktop computer in the lab.

Then, the second task was introduced by telling the participants that they would interact with a different virtual human. They were asked to put the HMD back on, and were being greeted by the second embodied agent ('Hi, my name is Alice [Alex]. Your task is to respond to ten questions as best as you can. I will be your advisor.'). The procedure was exactly the same as in the first task. Participants were asked what they thought the purpose of the study was. Finally, participants were debriefed and thanked for participating in the study.

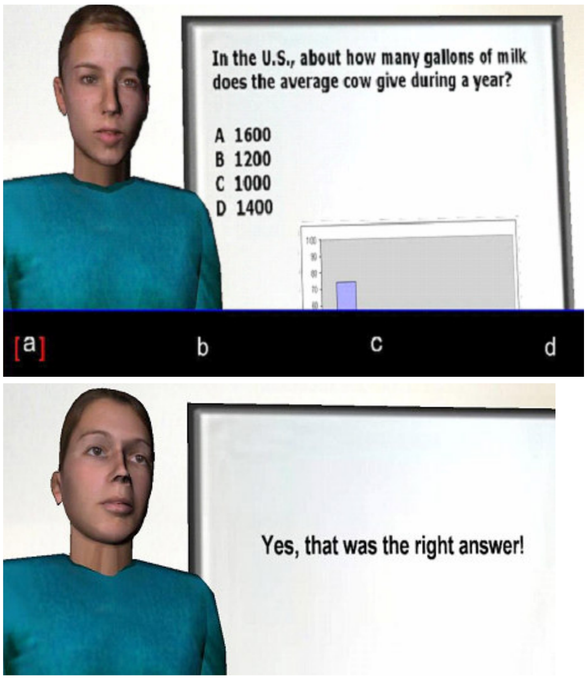


Figure 4.5: Answer selection (top), and answer feedback (bottom).

4.2.7 Measures

All measures were taken by means of a computer questionnaire containing construct-specific scales (as recommended by Krosnick and Fabrigar, 2007). Each item was followed by a 6-point rating scale, ranging from 1 (Not _ at all), 2 (Very little _), 3 (Somewhat _), 4 (Quite _), 5 (Very _), to 6 (Extremely _). Existing scales were used when possible in the construction of the structured user-perception questionnaire (e.g., scales used in Van Vugt et al., 2006, 2007b). When necessary, items were translated and modified to the purpose of the investigation. The questionnaire consisted of 44 items in total.

Reliability analyses (N = 64) were performed on each set of items concerning separate scales. Selection criteria were 1) an optimal contribution to Cronbach's alpha by showing little or no increase in the alpha level when the item was deleted, 2) a minimal inter-item correlation of .60, 3) an inter-item total correlation within a scale bigger than the correlation of each item with another scale (discriminant validity), and 4) a minimum of 2 items per scale. Items that failed on one or more of these criteria were not included in the measurement scales used in subsequent analyses.

We checked the affordance manipulations by means of a *perceived affordance* scale. In order to avoid directing the participant in an affirmative direction (see Dillman, 2000), two items were indicative (e.g., 'How knowledgeable do you think

Table 4.1: Discerning the factors *involvement*, *distance* and *use intentions* using a factor analyses in both the similar (left) and the dissimilar (right) condition.

	Similar condition			*Dissimilar condition*		
	Component 1	Component 2	Component 3	Component 1	Component 2	Component 3
Involvement1	.488	-.266	**.688**	.806	-.189	**.370**
Involvement2	.226	-.136	**.842**	.127	-.083	**.912**
Involvement3	.030	-.046	**.893**	.441	-.084	**.800**
Involvement4	.571	-.119	**.534**	.538	-.322	**.569**
Distance1	-.126	**.892**	-.160	-.164	**.839**	-.160
Distance2	-.444	**.820**	-.009	-.300	**.883**	.042
Distance3	-.640	**.514**	.187	-.277	**.843**	-.046
Distance4	-.212	**.833**	-.281	-.080	**.846**	-.248
UseIntention1	**.865**	-.222	.226	**.886**	-.226	.141
UseIntention2	**.785**	-.265	.419	**.869**	-.270	.285
UseIntention3	**.566**	-.430	.398	**.602**	-.606	.120

X is?') and two items were counter-indicative (e.g., 'How dumb do you think X is?'). The scale was reliable (Cronbach's alpha = .91).

We used a *perceived similarity* scale to determine whether or not participants detected. This scale consisted of two indicative items (e.g., 'How much do you think you and X look alike?') and two counter-indicative items (e.g., 'How much do you think X looks different than you?') and was reliable (Cronbach's alpha = .90).

One item from the *use intention* scale was discarded because of poor fit (criteria 1). The remaining scale consisted of two indicative items (e.g., 'How much do you want to use X again?') and one counter-indicative item ('How much do you want to get rid of X?'). The scale was reliable (Cronbach's alpha = .89).

Based on previous research, we assumed that *involvement* and *distance* are two distinct experiences that can occur at the same time. To demonstrate that involvement and distance should indeed be treated as different factors, and use intentions is a third separate factor, we performed a factor analyses with the involvement and distance used in the construction of these scales, and the three use intention items (varimax rotation, rotation converged in 6 iterations). Table 4.1 shows clearly that the three use intention items all loaded high on factor 1 (and, with one exception, low on factor 2 and 3), the four distance items all loaded high on factor 2 (and, with one exception, low on factor 1 and 3) and the four involvement items loaded high on factor 3 (and, with two exceptions, low on factor 1 and 2) in both the similar and the dissimilar condition.

In addition, we calculated the correlation between the involvement and the distance factor, which was r = -.42 in the dissimilar condition, and r = -.44 in the similar condition. These low correlations indicated that involvement and distance should be treated as separate dimensions (e.g., Neter et al., 1990). Both analyses justified treating the distance scale and the involvement scale as separate variables in subsequent analyses. The *involvement* scale consisted of four items (e.g., 'How much does X appeal to you?', 'How much do you feel connected to X?') and was reliable according to a Cronbach's alpha of .85. The *distance* scale consisted of four items (e.g., 'How much does X leave you with cold feelings?') and was reliable according to a Cronbach's alpha of .89.

Finally, the questionnaire measured the factors *perceived relevance* (3 items, Cronbach's alpha = .96) and *perceived valence* (4 items, Cronbach's alpha = .96) as potential mediators (as in Van Vugt et al., 2006), and included some questions regarding personal information about the participants such as gender, age, ethnicity, education, computer experience, game experience, and virtual reality experience.

4.3 Results

4.3.1 Morph detection

To ensure that our similarity manipulation was not obvious to the participants, we asked them to guess the intent of the experiment. Most participants guessed that the study was about different appearances of the agent, or about different voices. This was not surprising, as they indeed interacted with two agents with different appearances and voices in the two tasks. Six out of sixty-four participants suggested that the study could have been about facial similarity. In addition, a univariate analysis of variance (ANOVA), with designed similarity as the within subject factor and perceived similarity as the dependent variable, was conducted to evaluate whether participants perceived (or reported) the agent that was morphed with their own face as more similar to them than the agent that was morphed with a different face. The results indicated that *perceived* similarity ratings did not differ significantly between the similar and the dissimilar conditions, M(similar) = 2.34, SD = .77; M(dissimilar) = 2.30, SD = .87; $F(1,59) = .215$, $p = .65$, partial eta-squared = .004. Thus, the similarity manipulation was not obvious. The results of hypothesis testing did not change when we controlled for whether participants had detected the facial similarity manipulation or not.

4.3.2 Manipulation check affordances

We assessed the effectiveness of the affordance manipulations (aid versus obstacle, or, high versus low quality advice) in both the similar and the dissimilar condition by performing an ANOVA with designed affordances as the between-subject factor, designed similarity as the within-subject factor, and perceived affordance as the dependent variable. The aid and obstacle conditions had a significant effect on perceived affordances, $F(1,58) = 69.35$, $p < .001$, partial eta-squared = .95. Participants perceived the embodied agent in the aid condition as having better affordances than the embodied agent in the obstacle condition, both in the similar condition, M(aid) = 4.81, SD = .90; M(obstacle) = 2.91, SD = .86, and the dissimilar condition, M(aid) = 4.73, SD = 1.0; M(obstacle) = 2.95, SD = .67. Thus, we successfully manipulated affordances by means of varying the quality of an agent's advice.

4.3.3 Testing hypotheses

Testing hypotheses H1 and H2: direct effects of similarity and affordances on the dependent variables

To test the effect of one or multiple manipulated variables (designed affordances and designed similarity) in experimental situations have on a *single* dependent variable (e.g., involvement), we usually conduct a univariate analysis of variance (ANOVA). However, in the present study, we measured *multiple* dependent variables (involvement, distance, and use intentions) that could not easily be combined, so we performed a *multivariate* analysis of variance (MANOVA) (for more information, see Neter et al. (1990), or Lattin et al. (2003)). This can protect us against observing an effect where there is none (also called a Type I error), which might occur if we conducted multiple ANOVA's independently.

Yet, our design was a little more complex. Half of the participants performed the tasks with an agent that gave good advice, and the other half performed the tasks with an agent that gave bad advice, so affordances were manipulated between subjects. Furthermore, in one task participants interacted with their virtual self, and in the other task they interacted with a virtual other, so the similarity variable was manipulated within subjects. The between versus the within design prompted the use of a *mixed* design. In sum, because we have several dependent variables in our study (involvement, distance, and use intentions) and use a mixed design, we will perform *mixed MANOVA* to test the hypotheses. Lastly, we controlled for several user characteristics (e.g., age, computer experience) in the data analyses and we tested whether the gender of the participant affected the user responses by including gender as an additional factor in the analyses.

Thus, we performed a mixed MANOVA with designed similarity as the independent within factor, designed affordance and gender as the independent between factors, and involvement, distance, and use intentions as the dependent variables. The multivariate results revealed a main effect of designed affordance ($F(3, 58) = 37.61$, $p < .001$, partial eta-squared = .66). Furthermore, the multivariate results revealed an interaction effect of gender, designed similarity and designed affordance ($F(3, 58) = 3.21$, $p < .030$, partial eta-squared = .14). No other main or interaction effects reached significance (ps $> .10$). To interpret these multivariate results, that is, the effects of the independent variables on the *single* dependent variables, it is common practice to inspect the results of the univariate ANOVAs. Thus, we conducted ANOVAs on each dependent variable as follow-up tests to the MANOVA.

Use intentions. First, designed affordance had a significant effect on use intentions ($F(1, 60) = 107.27$, $p < .001$, partial eta-squared = .64). To interpret the direction of effects, we reported the means and standard deviations in Table 4.2. Participants were more eager to use an aiding agent than an obstructing agent again. This result fits general expectations and previous findings (Van Vugt et al., 2006). Second, gender, designed similarity and designed affordance interacted in

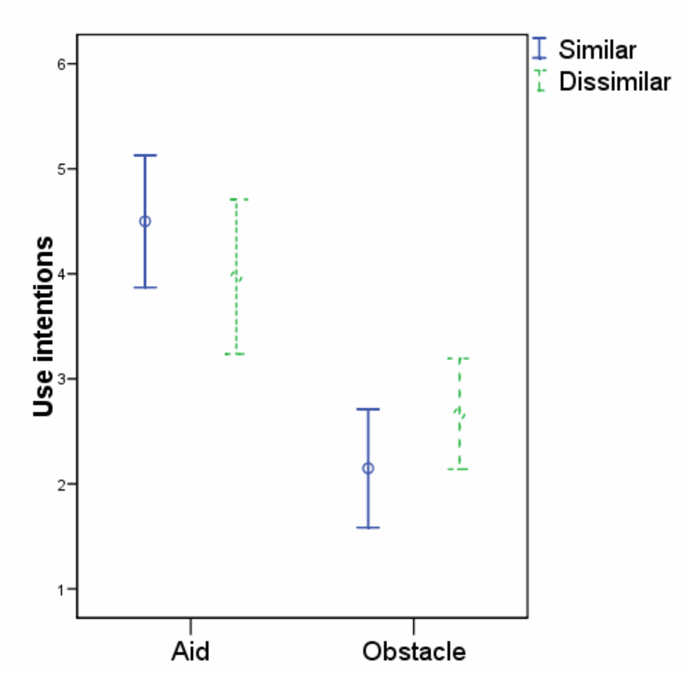


Figure 4.6: The effect of designed similarity and designed affordance on males' intentions to use the agent again.

explaining use intentions ($F(1, 60) = 4.176$, $p < .045$, partial eta-squared = .07). Females seemed unaffected by designed similarity (see Table 4.2). Males, however, were affected by designed similarity (see Figure 4.6). As expected (H1a), males were more willing to use an aiding agent when the agent was similar rather than dissimilar to them. Furthermore (H2a), males were less willing to use an obstructing agent that was similar rather than dissimilar to them. Thus, these results support H1a and H2a for males, but not for females.

Involvement. First, designed affordance affected involvement ($F(1, 60) = 17.77$, $p < .001$, partial eta-squared = .228). Participants felt more involved with an aiding agent than an obstructing agent, which fits general expectations (see Table 4.2). Second, the interaction effect of designed similarity and designed affordance on involvement was not significant, but the data did show a trend ($F(1, 60) = 3.65$, $p < .061$, partial eta-squared = .057). This indicates that, as expected (H1b), both male and female users were slightly more involved with an aiding agent when the agent was similar than when the agent was dissimilar to them. However, involvement remained unaffected by similarity when the agent was obstructing (see Table 4.2), contrary our expectations (H2b). Thus, we found support for H1b but not for H2b.

Distance. Designed affordance had a significant effect on distance ($F(1, 60) = 16.74$, $p < .001$, partial eta-squared = .218). Participants felt less distance from an

Table 4.2: Means and standard deviations of involvement (I), distance (D), and use intentions (U) in the conditions of designed affordance (aid versus obstacle), and designed similarity (similar versus dissimilar).

		Aid condition	*Obstacle* condition
Similar condition	Males	*I* 2.48 (1.05)	*I* 1.83 (.51)
		D 1.56 (.52)	*D* 2.72 (1.45)
		U 4.50 (.98)	*U* 2.14 (.78)
	Females	*I* 3.06 (.73)	*I* 1.87 (.47)
		D 1.52 (.45)	*D* 2.28 (.84)
		U 4.70 (.68)	*U* 2.38 (.66)
Dissimilar condition	Males	*I* 2.33 (1.10)	*I* 1.87 (.78)
		D 1.81 (.84)	*D* 2.40 (.74)
		U 3.97 (1.15)	*U* 2.66 (.73)
	Females	*I* 2.78 (.77)	*I* 2.00 (.60)
		D 1.93 (.89)	*D* 2.60 (.87)
		U 4.51 (1.16)	*U* 2.30 (.69)

aiding agent than an obstructing agent, which is in line with general expectations. No significant interaction effects were found on distance (ps $> .05$).

Thus, results show that the quality of affordances is important for involvement, distance, and use intentions. Similarity may also contribute to involvement and use intentions, but not always in a positive manner. Male users are more willing to use *dis*similar others than similar others if paired with low quality affordances. This highlights the importance of designing agents that have high quality affordances.

Testing hypotheses H3 and H4: *in*direct effects of facial similarity and affordances on the dependent variables

We wished to identify whether the variables similarity and affordances *in*directly affected the dependent variables (involvement, distance, and use intentions) via a third explanatory variable, perceived relevance or perceived valence (known as a mediator variable). Therefore, we performed *mediation* analyses. A mediation model hypothesizes that the independent variable affects the mediator variable, which in turn affects the dependent variable. We performed several mediation analyses to test whether designed similarity indirectly affected the dependent variables (H3), and whether designed affordances indirectly affected the dependent variables (H4). To test mediation, we used the Sobel method (Preacher and Hayes, 2004), which is reliable. If the Sobel test indicates a significant effect ($p < .05$), we may to conclude that mediation has occurred.

First, we used the Sobel method to test whether perceived relevance and perceived valence also mediated the relationship between designed similarity and the dependent variables of involvement, distance, and use intentions (H3). However, these tests suggested no mediation effect of perceived relevance or perceived valence between designed similarity and involvement, distance, or use intentions (ps

$> .10$). Thus, we did not find support for hypotheses H3a and H3b. In other words, designed similarity did not affect the dependents indirectly via perceptions of relevance and valence. Hence, in the present study, participants did not perceive the embodied agent to be more relevant for their goals, nor did participants have higher outcome expectations, when the agent was facially similar to them.

Second, we tested whether perceived relevance and valence mediated the relationship between affordances and (one of) the dependents (H4). In both the similar and dissimilar conditions, the effect of perceived affordances on *(1) use intentions* was mediated by both perceived relevance (similar: Sobel $z = 6.70$, $p < .001$; dissimilar: Sobel $z = 5.40$, $p < .001$) and perceived valence (similar: Sobel $z = 5.90$, $p < .001$; dissimilar: Sobel $z = 5.09$, $p < .001$). Thus, perceived affordances affected use intentions indirectly, through perceptions of relevance and valence. Second, the effect of perceived affordances on *(2) distance* was mediated by perceived valence in both the similar and dissimilar conditions (similar: Sobel $z = -2.01$, $p < .05$; dissimilar: Sobel $z = -2.68$, $p < .01$), and by perceived relevance in the dissimilar condition but not in the similar condition (similar: Sobel $z = -1.67$, $p = .09$; dissimilar: Sobel $z = -2.11$, $p < .04$). Here also, perceived affordances affected user distance towards the agent indirectly, through valence and, to a lesser extent, relevance. Third, results of the mediation tests revealed that the effect of perceived affordances on *(3) involvement* was *not* mediated by perceived relevance (similar: Sobel $z = 1.03$, $p = .30$; dissimilar: Sobel $z = .24$, $p = .81$) nor by perceived valence (similar: Sobel $z = .97$, $p = .33$; dissimilar: Sobel $z = -.14$, $p = .89$) in both the similar and dissimilar condition. In sum, similar to previous results (Van Vugt et al., 2006), perceived relevance and perceived valence were mediators between perceived affordances and use intentions, and between perceived affordance and distance, but not between perceived affordance and involvement. Thus, the higher the quality of advice (affordances), the more relevant the user perceives the embodied agent to be (perceived relevance) and the more positive outcomes the user expects (perceived valence). This leads to less user distance towards the agent and higher intentions to use the agent again. Thus, we found support for H4a, which concerned the mediating role of perceived relevance, and H4b, which concerned the mediating role of perceived valence.

Thus, results demonstrate the importance of user perceptions of relevance and valence for user responses to embodied agents. In addition, the results indicate that perceptions of relevance and valence are strongly affected by the affordances provided by an agent, but not by the agent's similar appearance.

4.4 Conclusion and discussion

The aim of the present study was to study the effects of facial similarity between agent and user under different conditions. More specifically, the agent had either helpful or obstructing affordances for the user. We expected that facial similarity would cause positive effects on user responses in case of an aiding agent, but negative effects in case of an obstructing agent. We also expected that effects of facial similarity and affordances on user responses would be indirect, via perceptions of relevance and valence. Measures were related to user affect (psychological involvement and distance) and to user behavior (use intentions).

Part of the expectations were supported by the data. As expected, both affordances and facial similarity affected user affect and user behavior. Indeed, facial similarity had positive effects on user responses in case of an aiding agent. However, that facial similarity would negative effects in case of an obstructing agent, was only true for the use intentions of males, not females. Furthermore, as expected, affordances on user responses were mediated through perceptions of relevance and valence. However, effects of facial similarity on user responses were *not* mediated through perceptions of relevance and valence.

The facial similarity effects that we found were remarkable because participants were *unaware* of the similarity manipulation. Ninety-one percent of the participants did not mention the facial similarity manipulation when asked about the purpose of the experiment, and perceived similarity ratings were equal among the designed similarity conditions. Thus, the facial similarity manipulation influenced participants' responses even though they did not explicitly detect it.

More specifically, the data revealed, first, an effect of facial similarity on involvement with an agent with aiding affordances. Both female and male users felt *more* involved with an embodied agent with *aiding* affordances when it was facially similar to them than when it was dissimilar. However, involvement with an embodied agent with *obstructing* affordances was equal for facially similar and facially dissimilar agents. Hence, facial similarity boosted involvement, but only when the agent was helpful for the user's task. Thus, in case of an obstructing agent, facial similarity did *not* evoke positive effects, but it did not cause negative effects either. Psychological distance remained unaffected by facial similarity and was merely depended on the agent's affordances - aiding affordances evoked lower psychological distance than obstructing affordances.

Second, the data revealed an effect of facial similarity on intentions to use the agent again in the future, but this occurred only for males. Female's intentions to use the agent again were mainly affected by the agent's affordances, aiding affordances evoked higher use intentions than obstructing affordances. Male's intentions to use the agent again were also affected by the agent's affordances, but also by facial similarity. If an embodied agent helped male users complete a task successfully (aiding affordances), facial similarity boosted males' use intentions. However, if an embodied agent kept male users from successfully completing the task (obstructing affordances), facial similarity was disadvantageous. In that case, males preferred to use a facial dissimilar agent. In previous research we found a comparable result, that is, if an agent was obstructing, *beauty* was disadvantageous for the use intentions of both male and female participants (Van Vugt et al., 2006).

Last, the data revealed that user perceptions of relevance and valence are important for user responses to embodied agents. In line with previous results (Van Vugt et al., 2006, 2007b), we found that perceptions of relevance and valence are affected by the affordances that an agent offers - an aiding agent was perceived as more relevant and users expected more positive outcomes from using an aiding agent (positive valence) than from using an obstructing agent. Consequently, users felt less distant with the agent and had higher intentions to use the agent again. Thus, relevance and valence served as mediators in between affordances and various user responses. However, perceptions of relevance and valence remained *un*affected by facial similarity.

Next, we will give several potential explanations for the unexpected differential effects of gender. After that, we will discuss the implications of the study for designers, indicate the limitations of the study, and present ideas for future research.

4.4.1 The differential effects of gender

We found three plausible explanations of why men, more than women, wanted to avoid interacting with an obstructing agent that looked like them. As we did not predict a gender difference, these explanations are necessarily ad hoc.

First, the gender difference might have been due to differences in *achievement needs* of men and women. Men are motivated by achievement needs and eminence to a greater extent than women (e.g., Hoffman, 1972; O'Neil, 1982). This is related to the male gender role that men have more agentic qualities such as independence and competitiveness than women and to the female gender role that women have more social qualities (Brody, 1997; Carli and Eagly, 1999). This also manifests itself in interaction with others. The communication style of males and females are based on different endeavors: whereas a man must strive to achieve independence and avoid failure, women strive to achieve closeness and consensus (Tannen, 1990). Thus, men want to be associated with success, and not with failure, whereas women find an association with success or failure less important. Therefore, men might feel an ego wound when interacting with an obstructing agent that looks like them, whereas women may not.

This links up with literature on gender and computer games indicating that women are less attracted to *competitive* elements in video games than men (Lucas and Sherry, 2004). Women care more for social interactions than for competition in computer games (Inkpen et al., 1994; Hartmann and Klimmt, 2006; Knobloch-Westerwick and Hastall, 2006). Perhaps in our studies, men wanted to win the 'competition' (score as high as possible on the questionnaires in the experiment), and did not want to be associated with failure. Women did not mind much to 'lose' and being associated with failure.

Second, technology-acceptance research states that men's technology-usage decisions are more strongly influenced by their *perceptions of usefulness* than those of women (Venkatesh, 2000). For men, it is more important than for women that technology is useful for their goals, thus, that an agent helps them to perform well. Consequently, being associated with a useless, obstructing agent, is something men may want to avoid more than women do.

Third, the gender differences might also have been due to differences in *task* *performance expectations*. Men tend to overestimate their performance more often than women do, and therefore, men might have approached the task with initially higher expectations of success than women (e.g., Hill and Dusek, 1969; Mura, 1987; Beyer, 1990; Beyer and Bowden, 1997). Men might have felt disappointed when they did not live up to their high expectations. Women, however, did not feel such disappointment because they had lower initial expectations. As a result, women did not mind whether the obstructing agent looked like them, whereas men did.

4.4.2 Implications for designers

Future embodied agent systems may monitor the user's physical appearance by using (web)cameras or photos stored on the computer. Then, the outer appearance of embodied agents can be made to resemble individual users automatically, similar to how web applications currently use 'cookies' (e.g., Millett et al., 2001). Morphing techniques could then blend two faces, for example the agent's face and the user's face, into one new face. These techniques allow the agent's face to resemble the user to varying degrees. Embodied agents can thus be made to resemble their user, possibly without the user's choice or awareness. In this paper, we will not discuss the potential ethical concerns surrounding such systems, which are likely to be comparable to the concerns surrounding cookies [ibid].

Such techniques are tempting to use. Facial similarity may positively affect involvement with an agent that is perceived by the user as an aiding tool. This is interesting for designers who aim to satisfy users, as involvement is an important predictor of user satisfaction (Lindgaard and Dudek, 2003; Van Vugt et al., 2006), and a prerequisite for better understanding (e.g., Falk and Dierking, 2000). Thus, facial similarity may be an interesting feature for embodied pedagogical agents that feature in computer assisted learning systems (e.g., Johnson et al., 2000; Moundridou and Virvou, 2002; Gulz and Haake, 2006; Person and Graesser, 2006; Aylett et al., 2007). Such systems may particularly profit from facial similarity when the users are male.

However, digital products fail to meet user goals with alarming frequency (e.g., Cooper et al., 2007). Thus, it is inevitable that some users become frustrated some of the time. When users find that embodied agents perform their task badly, facial similarity between agent and user may not affect involvement with the agent at all, and may even *decrease* the user's intentions to use the agent again. Designing for facial similarity seems a waste of time when the affordances are under-developed or, one way or the other, *perceived* by the user as obstructing. Thus, substantial effort should go into designing agents with high quality affordances that are relevant to user goals (cf. Norman, 1988; Preece et al., 2002). Avoid using self-representations in help assistants, advisors, or tutors that might be experienced as obstructing, especially when the computer users are expected to be male. It seems a better strategy to design agents with faces *dis*similar to the user's face than with similar or 'generic' faces. Only if users report that the agent is relevant for certain goals they have, and when they expect positive outcomes from using the application, may designers start to transform the embodied agent into an agent whose face looks like the user's, by morphing an image of the user with their original design. However, designers better not exceed a percentage of 35% facial similarity. After that, users become aware of the manipulation and may find the systems invades their privacy.

The above recommendations might especially apply to pedagogical and advisory agents like the ones we used in the experimental study, and might not be valid for all types of agents, and for all users. In certain types of applications, the visual appearance of an agent might be an important part of enjoyment and affect usage. For example, applications designed for entertainment purposes, and for certain types of users (e.g., children) might be more affected by looks than others. Thus, designers of agents that aim at, for example, entertaining children, may well put effort into visual appearance aspects such as facial similarity. However, even for such agents, high quality affordances should be preserved (cf. Cooper et al., 2007). In general, we advise that the design of embodied agents should not entirely exclude graphical concerns such as facial similarity, but rather should place such concerns within a functional framework (cf. Cooper et al., 2007).

4.4.3 Limitations and Future Work

A limitation of the present study is that it only examined same-gender interactions. Research showed several differences between same-gender and other-gender interactions. First, facial resemblance increases the attractiveness of same-gender faces more than other-gender faces (De Bruine, 2004a). Second, people have different expectations when the interact with a same-gender partner than with an opposite-gender partner (Rink and Ellemers, 2006). Third, in mixed-gender interactions, men and women behave in a less gender stereotypical manner than in same-gender interactions (cf. Carli, 1989; Guadagno and Cialdini, 2007). Because participants were in same-gender interactions, it might be that more gender stereotypical behavior emerged. Future work may investigate how males respond to female agents that look like them, and how females respond to male agents that look like them. Put differently, future research should investigate how facial similarity affects users in mixed-gender interactions, and examine whether our results can be generalized.

In addition, the combination of facial similarity and affordances might be investigated in different task contexts and settings. For example, the user's task may require working with agents in virtual teams in collaborative virtual environments. In such settings, it is probably important that these agents have a cooperative attitude. They should have similar interests as the user and show sufficient effort to fulfill the task successfully as a team. Based on our results, we expect that if agents are perceived as cooperative, facial similarity among team members may increase the user's involvement and use intentions. Perhaps, it may even increase the efficiency of doing the work. However, if agents are not perceived as cooperative, such agents probably better be facial dissimilar. Such research is important because collaborative virtual environments may become more prevalent in the future (e.g., Bailenson et al., 2005).

One may also study how facial similarity affects the user in more complex scenarios that are likely to exist in non-laboratory settings. In our experimental study, participants were instructed to focus on one particular goal, that is, participants had to answer the questions the best they could. However, multitasking is common in virtual environments (e.g., Cypher, 1986), and users may have multiple user goals, such as efficient task completion *and* enjoyment (e.g., Preece et al., 2002). Future research may inspect the effects of facial similarity and the differential effects of affordances and user gender in more complex settings.

In our study, we had a limited sample of participants due to practical reasons of feasibility. Future work should expand the sample size and test whether the effects we found are robust. In addition, we used only a single level of morphing to study facial similarity effects, and a morphing rate as low as 35% appeared enough to cause effects in 3D. We chose 35% because previous research indicated that this was the highest level a person's (2D) face could be morphed into another face without him or her explicitly noticing. However, the indiscernability of facial similarity might also explain why perceived relevance and perceived valence did not mediate facial similarity and user responses in the present study. The effects found in the present study might increase with the degree of facial similarity involved, which would be worth investigating. In follow-up studies, researchers should recognize that the potential of participant's awareness of the similarity manipulation might influence the results.

To conclude, it is not wise to believe that a user will use an embodied agent application more when the embodied agent looks like the user. Above all, designers should create a software application that is helpful as a tool. An advisor or virtual tutor should offer adequate information, so that the user perceives the tool as an aid to the task. If designers successfully do so, they will increase the user's personal involvement with the helpful agent. Then, they may even boost the enthusiasm to use the tool by making this helpful teacher, advisor, or coach look like its user.

4.5 Acknowledgments

This paper is the result of a collaborative research effort between the VU University Amsterdam (The Netherlands) and Stanford University (Palo Alto, USA). We would like to thank both universities for the collaboration opportunity. Further, we are grateful to the students that worked at the VHIL lab on this project as well as Nick Yee for their assistance. We are indebted to the editor, Marti Hearst, and three anonymous reviewers whose insightful comments and suggestions significantly enhanced the quality of this paper. We would also like to thank Gerrit van der Veer and Jan Kleinnijenhuis for their support and useful comments on earlier drafts of this paper.

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211.

Ames, D. R. (2004). Inside the mind readers toolkit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, 87:340353.

Aylett, R. S., Paiva, A., Woods, S., Hall, L., and Zoll, C. (2007). Expressive characters in anti-bullying education. In *Animating expressive characters for social interaction*. John Benjamins. In press.

Bailenson, J. N., Beall, A. C., Blascovich, J., and Rex, C. (2004). Examining virtual busts: Are photogrammetrically-generated head models effective for person identification? *PRESENCE: Teleoperators and Virtual Environments*, 13(4):416–427.

Bailenson, J. N., Iyengar, S., Yee, N., and Collins, N. (2008). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly*. In press.

Bailenson, J. N., Rex, C., Beall, A. C., and Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence*, 10(6):583–598.

Bailenson, J. N., Swinth, K. R., Hoyt, C. L., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4).

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research - conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5):960–970.

Beyer, S. and Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23:157–172.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., and Bailenson, J. (2002). Immersive virtual environment technology as a research tool for social psychology. *Psychological Inquiry*, 13:103–125.

Blumler, J. and Katz, E. (1974). *The uses of mass communications*. Sage Publications, Beverly Hills, CA.

Brehm, J. W. and Miron, A. M. (2006). Can the simultaneous experience of opposing emotions really occur? *Motivation and Emotion*, 30(1):13–30.

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology*, 1:650–654.

Brody, L. R. (1997). Gender and emotion: Beyond stereotypes. *Journal of Social Issues*, 53:369–394.

Byrne, D. (1971). *The attraction paradigm*. Academic Press, New York.

Cacioppo, J. T. and Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115:401–423.

Card, S., Moran, T., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Carli, L. L. (1989). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology*, 56(4):565–576.

Carli, L. L. and Eagly, A. H. (1999). Gender effects on social influence and emergent leadership. In Powell, G., editor, *Handbook of gender and work*, page 203280, Thousand Oaks, CA. Sage.

Cassell, J., Bickmore, T., Vilhjálmsson, H., and Yan, H. (2000). More than just a pretty face: affordances of embodiment. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 52–59. ACM Press.

Catrambone, R., Stasko, J., and Xiao, J. (2004). ECA as user interface paradigm. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 239–270, Dordrecht / Boston / London. Kluwer.

Chaiken, S. (1979). Communicator attractiveness and persuasion. *Journal of Personality and Social Psychology*, 37:1387–1397.

Cialdini, R. B. (2001). *Influence: Science and practice*. Harper Collins, New York, 4th edition.

Cooper, A., Reimann, R., and Cronin, D. (2007). *About Face 3: The Essentials of Interaction Design*. John Wiley & Sons, Indianapolis, Indiana, USA.

Cypher, A. (1986). The structure of user's activities. In Norman, D. A. and Draper, S., editors, *User centred systems design*, pages 243–263, Hillsdale. Lawrence Erlbaum Associates, Inc.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–339.

De Bruine, L. M. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society London B*, 269:13071312.

De Bruine, L. M. (2004a). Facial resemblance increases the attractiveness of same-sex faces more than other-sex faces. *Proceedings of the Royal Society London B*, 271:20852090.

De Bruine, L. M. (2004b). Resemblance to self increases the appeal of child faces to both men and women. *Evolution and Human Behavior*, 25:142–154.

Dehn, D. M. and Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22.

Diener, E. and Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, 47(5):1105–1117.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. John Wiley and Sons, New York, 2th edition.

Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence*, 13(3):273–295.

Eidelman, S. and Biernat, M. (2003). Derogating black sheep: Individual or group protection? *Journal of Experimental Social Psychology*, 39:602–609.

Falk, J. H. and Dierking, L. D. (2000). *Learning from museums, visitor experiences and the making of meaning*. AltaMira Press, Walnut Creek, CA.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7:117–140.

Fishbein, M. and Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley, Reading, MA.

Frijda, N. (1988). The laws of emotion. *American Psychologist*, 43(5):349–58.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press, Cambridge.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Goethals, G. R. and Darley, J. (1977). Social comparison theory: an attributional approach. In Suls, J. M. and Miller, R. L., editors, *Social comparison processes: Theoretical and empirical perspectives*, pages 86–109. Hemisphere.

Gollwitzer, P. M. and Bargh, J. A. (1996). *The psychology of action: Linking cognition and motivation to behavior*. Guilford, New York.

Gratch, J. and Marsella, S. (2005). Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, 19(3-4):215–233.

Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors (2006). *Proceedings of Intelligent Virtual Agents 2006, IVA'06, Marina del Rey, CA, USA, August 21-23, 2006*, volume 4133 of *Lecture Notes in Artificial Intelligence*. Springer.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. (2007). Virtual humans and persuasion: the effects of agency and behavioural realism. *Media Psychology*, 10:1–22.

Guadagno, R. E. and Cialdini, R. B. (2007). Persuade him by email, but see her in person: Online persuasion revisited. *Computers in Human Behavior*, 23:999–1015.

Gulz, A. and Haake, M. (2006). Design of animated pedagogical agents - a look at their look. *International Journal of Man-Machine Studies*, 64(4):322–339.

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9):602–615.

Hartmann, T. and Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication*, 11(4):article 2.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology*, 21:107–112.

Hersh, W. (1994). Relevance and retrieval evaluation: perspectives from medicine. *Journal of the American Society for Information Science*, 45(3):201–206.

Hill, K. T. and Dusek, J. B. (1969). Children's achievement expectations as a function of social reinforcement, sex of s, and test anxiety. *ChiM Development*, 40:547–557.

Hinsz, V. B. (1989). Facial resemblance in engaged and married couples. *Journal of Social and Personal Relationships*, 6:223–229.

Hoffman, L. W. (1972). Early childhood experiences and women's achievement motives. *Journal of Social Issues*, 28:129–155.

House, W. C. and Perney, V. (1974). Valence of expected and unexpected outcomes as a function of locus of goal and type of expectancy. *Journal of Personality and Social Psychology*, 29:454463.

Inkpen, K., Upitis, R., Klawe, M., Lawry, J., Anderson, A., Ndunda, M., Sedighian, K., Leroux, S., and Hsu, D. (1994). We have never forgetful flowers in our garden: Girls' responses to electronic games. *Journal of Computers in Math and Science Teaching*, 13(4):383–403.

Johnson, W. L., Rickel, J., , and Lester, J. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78.

Kaptelinin, V. and Nardi, B. A. (2006). *Acting with technology. Activity theory and interaction design theory*. MIT Press, Cambridge, MA.

Klohnen, E. C. and Luo, S. (2003). Interpersonal attraction and personality: what is attractive–self similarity, ideal similarity, complementarity or attachment security? *Journal of Social and Personality Psychology*, 85(4):709–722.

Knobloch-Westerwick, S. and Hastall, M. R. (2006). Social comparisons with news personae. Selective exposure to news portrayals of same-sex and same-age characters. *Communication Research*, 33(4):262–284.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A. and Van Vugt, H. C. (2008). Emotions in mediated interpersonal communication: Toward modeling emotion in virtual humans. In *Mediated interpersonal communication*, pages 100–130. Lawrence Erlbaum Associates/ Taylor and Francis/ Routledge.

Krosnick, J. A. and Fabrigar, L. R. (2007). *The handbook of questionnaire design*. Oxford University Press, New York. Forthcoming.

Kruglanski, A. W. and Mayseless, O. (1990). Classic and current social comparison research: expanding the perspective. *Psychological bulletin*, 108(2):195–208.

Langlois, J. H., Ritter, J. M., Roggman, L. A., and Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27:79–84.

Lanier, J. (2001). Virtually there. *Scientific American*, April:66–75.

Lattin, J., Carroll, J. D., and Green, P. E. (2003). *Analyzing multivariate data*. Duxbury.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press, New York.

Lee, E.-J. (2003). Effects of "gender" of the computer on informational social influence: the moderating role of task type. *International Journal of Human-Computer Studies*, 58(4):347–362.

Lee, E.-J. (2007). Wired for gender: Experientiality and gender-stereotyping in computer-mediated communication. *Media Psychology*, 10(2):182–210.

Leontiev, A. (1978). *Activity, consciousness, and personality*. Prentice-Hall, Englewood Cliffs, N.J. Original work published in Russian in 1975.

Lerner, M. J. and Agar, E. (1972). The consequence of perceived similarity: Attraction and rejection, approach and avoidance. *Journal of Experimental Research on Personality*, 6:69–75.

Levinger, G. and Breedlove, J. (1966). Interpersonal attraction and agreement: a study of marriage partners. *Journal of Personality and Social Psychology*, 3:367–372.

Li, I., Forlizzi, J., Dey, A., and Kiesler, S. (2007). My agent as myself or another: effects on credibility and listening to advice. In *DPPI '07: Proceedings of the 2007 conference on designing pleasurable products and interfaces*, pages 194–208, New York, NY, USA. ACM.

Lindgaard, G. and Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15(3):429–452.

Lucas, K. and Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication Research*, 31:499–523.

McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface 2000*, pages 179–186.

Miller, N. and Marks, G. (1982). Assumed similarity between self and other: Effect of expectation of future interaction with that other. *Social Psychology Quarterly*, 45:100–105.

Millett, L. I., Friedman, B., and Felten, E. (2001). Cookies and web browser design: toward realizing informed consent online. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 46–52, New York, NY, USA. ACM.

Moundridou, M. and Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning*, 18(3):253–261.

Mumford, M. D. (1983). Social comparison theory and the evaluation of peer evaluations: a review and some applied implications. *Personnel Psychology*, 36(4):867–881.

Mura, R. (1987). Sex-related differences in expectations of success in undergraduate mathematics. *Journal for Research in Mathematic Education*, 18(1):15–24.

Nass, C., Kim, E.-Y., and Lee, E.-J. (1998). When my face is the interface: an experimental comparison of interacting with one's own face or someone else's face. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 148–154, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103.

Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied linear statistical models*. Irwin, Homewood, IL, 3rd edition.

Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.

Nowak, K. L. and Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1):article 8.

Olson, I. R. and Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, 5(4):498–502.

O'Neil, J. M. (1982). Gender-role conflict and strain in men's lives: implications for psychiatrists, psychologists, and other human-service providers. In Solomon, K. and Levy, N. B., editors, *Men in Transition: Theory and Therapy*, pages 5–44, New York. Plenum.

Payne, C. and Jaffe, K. (2005). Self seeks like: Many humans choose their dog-pets following rules used for assortative mating. *Journal of Ethology*, 23:15–18.

Person, N. P. and Graesser, A. C. (2006). Pedagogical agents and tutors. In Guthrie, J. W., editor, *Encyclopedia of Education*, pages 1169–1172, New York. Macmillan.

Pratt, J. A., Hauser, K., Ugray, Z., and Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers*, 19:512–523.

Preacher, K. J. and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4):717–731.

Preece, J., Rogers, Y., and Sharp, H. (2002). *Interaction design: beyond human-computer interaction*. John Wiley and Sons, Inc.

Prendinger, H. and Ishizuka, M., editors (2004). *Life-like characters. Tools, affective functions, and applications*. Cognitive Technologies Series. Springer.

Priester, J. R. and Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: Interpersonal and intrapersonal antecendents of evaluative tension. *Journal of Personality and Social Psychology*, 80:19–34.

Ratan, R. and Bailenson, J. N. (2007). Similarity and persuasion in immersive virtual reality. Panel presentation to the Communication and Technology Commission of ICA.

Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge University Press.

Rink, F. and Ellemers, N. (2006). What can you expect? the influence of gender diversity in dyads on work goal expectancies and subsequent work commitment. *Group Processes and Intergroup Relations*, 9(4):577–588.

Russell, J. A. and Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, 125(1):3–30.

Ruttkay, Z., Dormann, C., and Noot, H. (2004). Embodied conversational agents on a common ground. A framework for design and evaluation. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 27–66. Kluwer Academic Publishers.

Ruttkay, Z. and Pelachaud, C. (2004). *From brows to trust: Evaluating embodied conversational agents*. Kluwer Academic Publishers, Dordrecht / Boston / London.

Saffer, D. (2007). *Designing for interaction. Creating smart applications and clever devices*. New Riders.

Sangrador, J. L. and Yela, C. (2000). What is beautiful is loved: physical attractiveness in love relationships on a representative sample. *Social Behavior and Personality*, 28(3):207–218.

Sergent, J. and Signoret, J.-L. (1992). Functional and anatomical decomposition of face processing: Evidence from prosopagnosia and PET study of normal subjects. *Philosophical Transactions of the Royal Society London Biological*, 335:55–62.

Silvia, P. J., Graham, J. S., and Hawley, C. N. (2005). Changing attitudes towards prison reform: effects of similarity to prisoners on attraction and rejection. *Journal of Applied Social Psychology*, 35(2):248–258.

Tannen, D. (1990). *You just don't understand: Men and women in conversation*. Ballantine Books, New York.

Taylor, S. E. and Mettee, D. R. (1971). When similarity breeds contempt. *Journal of Personality and Social Psychology*, 20(1):75–81.

Tesser, A. (1988). Towards a self-evaluation maintenance model of social behavior. In Berkowitz, L., editor, *Advances in experimental social psychology*, volume 21, pages 181–227, San Diego. Academic Press.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I., and Eliëns, A. (2007). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers*, 19(2):267–280.

Van Welie, M. and Van der Veer, G. C. (2004). Groupware task analysis. In Hollnagel, E., editor, *Handbook of cognitive task design*, pages 447–476, New Jersey, US. Lawrence Erlbaum Associates, Inc. In press.

Venkatesh, V. (2000). Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4):342–365.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

Yee, N. and Bailenson, J. N. (2007). The Proteus effect: Self transformations in virtual reality. *Human Communication Research*, 33:271290.

Yee, N., Bailenson, J. N., and Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, pages 1–10, New York, NY, USA. ACM Press.

CHAPTER 5

When too heavy is just fine: Creating trustworthy e-health advisors

Abstract

People compare embodied agents not only with their *actual* selves but with their *ideal* selves as well. In this paper, a laboratory and an online study scrutinized the effects of similarity with and idealness of an embodied agent feature on user involvement with, distance towards, and intentions to use the e-health advisor. The advisor's body size was either similar or dissimilar to the user's actual body size, and had an ideal (slim) or non-ideal (heavier) shape. Results indicated that idealness was more important than similarity for explaining involvement with, distance towards, and intentions to use the embodied agent, but in unexpected ways. The heavier e-health advisors were perceived as the more trustworthy ones. As a result, the heavier e-health advisors were more involving, less distancing, and evoked higher use intentions than slimmer e-health advisors. Sometimes, it is better to forget the stereotypical preference and design embodied agents that are not ideal.

This chapter is submitted for publication:
Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. (submitted). When too heavy is just fine: Creating trustworthy e-health advisors.

The first study presented in this chapter is published as:
Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. (2006) Why fat interface characters are better e-health advisors. In: Gratch, J. et al (Eds.), *Proceedings of Intelligent Virtual Agents 2006*, volume 4133 of Lecture Notes in Artificial Intelligence, pages 1-13. Springer.

5.1 Introduction

'You probably think I've got a nerve putting you on a diet' said the overweight doctor to his overweight patient. This cartoon at CSL CartoonStock (Parker, 2007) illustrates the peculiar situation of consulting a doctor who does not appear to live by his or her own rules. Because the doctor is as heavy as the patient is, the credibility of the doctor's advice may be somewhat unpromising. What can one expect from someone who does not exemplify the ideally sized role model? In real life, however, other cues tell you to trust the doctor, despite his appearance. A diploma on the wall suggests that the doctor has taken the Hippocratic oath, which means he supposedly is an ethical person. In other words, the doctor's being overweight, may give the patient negative expectations regarding the doctor's health advice; however, contextual cues that indicate the doctor's moral fiber may yield positive expectations simultaneously. The mixed conclusion suggests that you should not do what the doctor does but do what he says that is, follow his recommendation to diet. In addition, the doctor's obvious lack of success in losing weight himself, might might motivate the patient to lose more weight than him.

People often compare themselves to others on dimensions such as outer appearance (Festinger, 1954), which may affect user perceptions of similarity. A range of studies on embodied agents, such as e-health advisors, have focused on the effects of similarity between the agent and the user on the user's attitudes towards the agent. Overall these studies have found that users hold more positive attitudes towards embodied agents that are similar than to agents that are dissimilar to themselves (Nowak and Rauh, 2005; Guadagno et al., 2007; Bailenson et al., 2001; Nass and Moon, 2000; Dryer, 1999). Thus, research indicates that similarity attracts. In general, the advice is to design embodied agents that users feel emotionally involved with and not too distant from, and that they want to use in the future (Nowak and Rauh, 2005; Van Vugt et al., 2006). Given the literature on embodied agents, similarity may promote user involvement and use intentions, and reduce distance. Thus, H1 is formulated as:

H1 An embodied agent with a similar body size as the user evokes

(a) higher use intentions, and

(b) higher involvement / lower distance

than an embodied agent with a dissimilar body size.

Hence, H1 states that agents with a body size similar to the user's will evoke more positive responses than dissimilar agents. For example, thin users will feel more involved with thin agents giving health advice, and heavier users will feel more involved with heavier agents that do so.

However, the development of attitudes toward embodied agents involves more than similarity and dissimilarity alone. In Western society, thin figures are often perceived as 'better' or more ideal than heavier ones. That is, people consistently rate slim and slender figures as more beautiful than heavier figures and, in the media, slim people are overrepresented as attractive characters (e.g., Greenberg et al., 2003). People also tend to attribute more negative characteristics, such as laziness, lack of will power, and sloppiness, to overweight people than to slim people (e.g., Puhl and Brownell, 2001). Even doctors who treat obese patients have negative attitudes towards heavier people and adopt the Western anti-fat bias (Teachman and Brownell, 2001). People with an ideal thin body are often made an example. Supposedly, they give others hope and inspiration to become just like them. An individual's self image can be enhanced by socially desirable associations, and similarity to the ideal self can have positive effects for self-evaluation. In the literature, this is known as 'wishful identification' (e.g., Hoffner and Buchanan, 2005; Konijn et al., 2007) and 'role modeling' (Bandura, 2001). Whether through diet, exercise or plastic surgery, many people feel it is necessary to strive for this ideal appearance. In other words, in addition to similarity to the actual self, similarity to the ideal self may be an important predictor of attraction. Indeed, studies examining similarity to the actual self versus similarity to the ideal self suggest that, in real life, similarity to the ideal self is not only an important predictor of attraction; at times it is even *more* important than similarity to the actual self (Wetzel and Insko, 1982; LaPrelle et al., 1990). Despite the seeming importance of similarity to the ideal, the concept has remained relatively unexplored in embodied agent research. Therefore, the present study addresses ideal similarity in contrast to actual similarity. This poses the interesting question of whether ideal similarity will prevail over actual similarity in predicting involvement with, distance towards, and intentions to use an e-health advisor.

Generally, people do not want to see their non-ideal features reflected in others (cf. Van Vugt et al., 2007a). Under certain circumstances, however, people do not want to compare themselves to ideal others either, because this contrasts a person's 'self' against the more successful other (e.g., Tesser, 1988). In cases of upward social comparison, non-ideal heavier people might feel threatened by an ideally thin advisor who seems better than them. Therefore, heavier people might prefer to avoid talking to a thin advisor and wish to talk with a heavier advisor instead. In cases of downward social comparison, ideal, slim people may wish to compare themselves with non-ideal, heavier others who are worse off, because this makes them feel good about themselves (e.g., Wills, 1981; Suls et al., 2002). Thus, ideally thin people might feel better talking to a non-ideal, heavier advisor than to an advisor that is just as ideal as they are. Thus, both people with a non-ideal body size and those with an ideal body size might feel more involved with a non-ideal advisor than with an ideally thin one, but for different reasons.

Yet, such non-stereotypical effects exist only when the dimension of comparison is *important*, or relevant, for self-definition (Bers and Rodin, 1984; Tesser, 1986, 1988; Salovey and Rodin, 1991). Thus, assuming that body size is an important dimension of self-definition in a health context, people of all body sizes may prefer non-ideal heavier e-health advisors over ideal ones, though for different reasons. Because food intake and diet are related to body size (cf. Harrison et al., 2006), and because overweight and obesity as well as anorexia lead to serious health consequences (e.g., World Health Organisation, 2000, 2003), we can regard body size as an important feature in health contexts. Thus, given the literature on similarity to an ideal, H2 challenges H1 in stating that:

> **H2** If the dimension of comparison between self and embodied agent is important in the context, then a *non*-ideal embodied agent evokes
>
> (a) higher use intentions, and
>
> (b) higher involvement / lower distance
>
> than an ideal e-health advisor.

Thus, in H1, based on the similarity literature, we expect that people will be more willing to use, more involved with, and less distant towards e-health advisors with a body size similar to their own. However, in H2, based on the literature on ideal similarity, we expect that people (of all sizes) are more willing to use, more involved with, and less distant towards non-ideal e-health advisors.

But this is still not the full story. A person's willingness to take health advice depends on the perceived trustworthiness and credibility of the advisor. Trust is of utmost importance in real-life health contexts. Trust and related socio-emotional aspects of the doctor-patient relationship have been associated with adherence to prescribed medication use (Piette et al., 2005), patient health outcomes (Stewart, 1995), patients' reported satisfaction with health care (DiMatteo and Hays, 1980), and patients' evaluation of their health care experience (Calnan and Rowe, 2006). Thus, when someone trusts their doctor, the doctor's treatment is more likely to be effective. Nowadays, 26% of teenage Internet users and 57% of adult Internet users look for health information online (Lenhart et al., 2001). Because so many users obtain health information from the Internet, it is important that users know how to trust the information. Therefore, Web sites increasingly use an organizational domain, display a copyright, and have other indicators of accuracy such as a HON code logo[1]. Moreover, assuming that health information proliferating on the Internet is accurate, the credibility of the advising agent is also very important, especially because patients generally interact with an e-health advisor or care giver on personal issues. Thus, when users encounter an e-health advisor on the screen, they are likely to not only compare themselves to the agents, but also assess whether or not the e-health advisor can be trusted. Such perceptions may subsequently affect the user's involvement with, distance towards, and intentions to use the e-health advisor. Perceptions of ethics refers to people's innate response to judge others in terms of 'good' versus 'bad', cf. 'trustworthy' versus 'not trustworthy'. Previous work on viewers' responses to fictional characters, such as those in films, found that the perception of ethics was *the* most important predictor of viewers' emotional involvement with, and emotional distance towards, such characters (Konijn and Hoorn, 2005). Would this also apply to e-health advisors?

[1]http://www.hon.ch/index.html

Real-life doctors have medical diplomas and take the Hippocratic oath, and supposedly adhere to high moral standards. In addition, the outer appearance of a real-life doctor also affects trust. People favor doctors in professional attire that includes a white coat and a name badge (e.g., Kanzler and Gorsulowsky, 2002; Rehman et al., 2005). E-health advisors, however, do not have medical degrees (nor do their designers) and do not take the Hippocratic oath (cf., 'I'm not a real doctor, but I play one in virtual reality', Shapiro and McDonald, 1992). Thus, it is difficult for e-health seekers to trust e-health advisors based on their expertise and adherence to moral standards. Because of this, the outer appearance of an e-health advisor may be the only way for users to estimate trust and credibility. Therefore, we wonder how users' perceptions of ethics are related to an e-health advisor's body size. Does the stereotypical perception that 'what is beautiful is good' (Dion et al., 1972) prevail in an e-health context or not? Therefore, we pose the following research question:

RQ1 Does the body size of an e-health advisor affect perceptions of ethics, which subsequently affect involvement, distance, and use intentions?

Thus, in this paper we argue that the body size of an e-health advisor may affect the user in several ways. First, the body size of an e-health advisor may evoke comparison processes, affecting user involvement with, distance towards and intentions to use e-health advisors (as formulated in H1 and H2). Second, the body size of an e-health advisor may affect users' perceptions of ethics - is the doctor 'good' or 'bad'. Perceptions of ethics are also likely to affect user involvement with, distance towards and intentions to use e-health advisors (cf. Bickmore and Cassell, 2001; Bickmore et al., 2005; Van Vugt et al., 2006). It is not yet clear which of the two, comparison processes or perceptions of ethics, is more important for understanding the various user responses to e-health advisors. Therefore, we posed the following research question as well:

RQ2 Which processes will better explain intentions to use, involvement with, and distance towards e-health advisors, comparison processes or perceptions of ethics?

To test the various hypotheses and to address the research questions, we conducted two studies. The first study was performed in the laboratory of a Western university with university students, whereas the second study addresses a more general public with a greater diversity in body weight of participants.

5.2 Study 1

5.2.1 Method

Design

A 2 (similarity: similar versus dissimilar) x 2 (idealness: ideal versus non-ideal) factorial design was used to test the various hypotheses, and to inspect the research questions (see Table 5.1). Assignment of participants to experimental conditions was slightly unbalanced because the condition ‘similar and ideal’ (see Table 5.1), could only be created after data-collection, because it depended on the number of participants indicating the exact *same* figure to represent both their actual and their ideal body size.

Table 5.1: Number of participants of Study 1 (*n*), means (*M*) and standard deviations (*SD*) of the body size of the used e-health advisors in the similarity and idealness conditions, based on the 1-9 range of the Figure Rating Scale.

	Similar to actual self	**Dissimilar to actual self**
Ideal	similar and ideal	dissimilar but ideal
	n = 17	n = 14
	M(size) = 3.3, SD = .92	M(size) = 3.5, SD = .66
Non-ideal	similar but non-ideal	dissimilar and non-ideal
	n = 18	n = 27
	M(size) = 4.2, SD = 1.4	M(size) = 7.7, SD = 2.9

Participants

Participants in our experiment were 80 university students (24 males and 56 females; age M = 23, SD = 7.8). Most Body Mass Index ratings could be categorized as normal according to the classification of the World Health Organization (BMI M = 22 kg/m^2, SD = 3.8). The students were paid 2.50 Euro for their volunteered participation.

Materials

An online, colored, and modernized version of the standardized Figure Rating Scale (FRS)[2], see Figure 5.1, was used to measure actual, ideal, and non-ideal body sizes, as perceived by the participant. Male participants were shown the male version (upper row) and female participants were shown the female version (lower row). The Figure Rating Scale is considered to be a reliable measure that is highly related to the Body Mass Index (Stunkard et al., 1983; Fingeret et al., 2004; Bulik et al., 2001).

One of these figures was used as the embodied agent with which the participant interacted in the e-health context. Participants were randomly assigned to interact

[2]after Stunkard et al. (1983), with permission

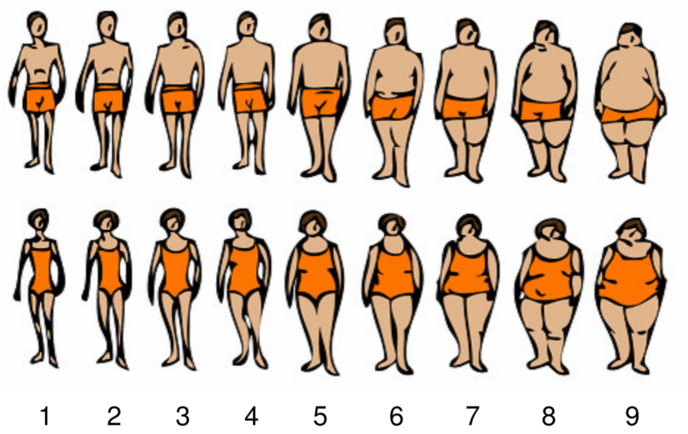


Figure 5.1: Modernized version of the Figure Rating Scale.

with an e-health advisor that was equal to their actual-self, to their ideal self, or to their non-ideal self. Table 5.1 shows the means and standard deviations of the body size of the used e-health advisors in all conditions. The body size numbers refer to the 1-9 range of the Figure Rating Scale. The e-health advisor always had the same gender as the participant, as gender may influence similarity perceptions (e.g., Nowak and Rauh, 2005; Guadagno et al., 2007). The embodied agent was called René (male) or Renée (female) and was positioned centrally on a Web site and enlarged to occupy a large part of the computer screen (see Figure 5.2). René(e) had four different poses and the text was positioned right next to him/her.

On the Web site, René(e) asked participants their opinion or knowledge on several health-related issues, using closed-answered questions, sometimes preceded by small introductory texts. For example, ‘Three quarters of Internet users, about 9 million people, search for information on health issues on the Internet. Do you

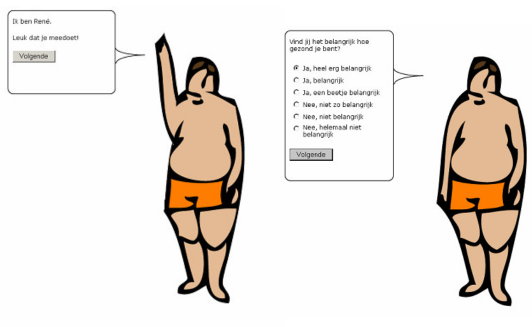


Figure 5.2: Screenshots of the René(e) software.

search the Internet for health information?', or, 'How important is your health to you?'. At the end, René(e) told that 'Soon, you can ask me questions on the Internet about a healthy lifestyle! For example, do you eat healthy?' We used questions, as opposed to solely plain text, to ensure a relatively lively interaction between the embodied agent and the participant, and to get more insight into their (un)healthy behaviors[3].

Procedure

Participants were seated individually, in front of a computer screen in one of the cubicles in a research lab at our university. The participants were welcomed by the computer and told that their data would be processed anonymously. Then, they were asked to enter their gender, age, weight, and height. After that, they were asked to rate their actual-self (which figure do you resemble most), their ideal-self (which figure do you want to look like), and their non-ideal-self (which figure do you not want to look like) on three Figure Rating Scales, on separate Web pages. The three scales were presented randomly to prevent order effects. The computer did not allow unanswered questions. Next, the figure selected automatically from the Figure Rating Scale appeared on the screen in the form of an embodied agent (i.e. an e-health advisor). On the first page, the embodied agent asked for the participant's name, then introduced itself as either René (male) or Renée (female), and welcomed the participant. In the following pages, René(e) gathered personal information and asked participants their opinion or knowledge on several health-related issues (as explained in section 5.2.1).

Next, the automatically selected figure of the Figure Rating Scale appeared on the screen in the form of an embodied agent (i.e. an e-health advisor). At the first page, the embodied agent asked for the participant's name, then introduced itself as either René (male) or Renée (female), and welcomed the participant. In the following pages, personal information was gathered and René(e) asked participants their opinion or knowledge on several health-related issues (as explained in section 5.2.1). After the interaction, which took about 6 minutes, the participant was asked to complete a user perception questionnaire, presented on several subsequent Web pages. After completing the questionnaire, participants were debriefed and dismissed.

Measurements

All measurements were taken by means of a questionnaire containing Likert-type scales. Each item was followed by a 6-point rating scale, ranging from 1 (do not agree at all), 2 (do not agree), 3 (barely agree), 4 (agree a little), 5 (agree), to 6 (fully agree). Items were presented in random order. For the present study, we used shortened versions of reliable scales used in previous experiments (Van Vugt et al., 2006, 2007b). Where necessary, items were adjusted to the purpose of the

[3]These results will be addressed elsewhere.

present study, the specific materials, and the language use of the target group of participants (university students).

Reliability analyses (N = 80) were performed on each set of items concerning separate scales. Selection criteria were 1) an optimal contribution to Cronbach's alpha by showing little or no increase in the alpha level when the item was deleted, 2) a minimal inter-item correlation of .60, 3) an inter-item total correlation within a scale bigger than the correlation of each item with another scale (discriminant validity), and 4) a minimum of 2 items per scale. Items that failed on one or more of these criteria were not included in the measurement scales used in subsequent analyses.

Involvement and distance. Involvement and distance were measured using 3 items each, based on previous work (e.g., Konijn and Hoorn, 2005; Van Vugt et al., 2006). The involvement items were: 'Do you feel good about me?', 'Do you feel involved with me', and 'Do you think it is pleasant to deal with me?'. The scale was reliable with a Cronbach's alpha of .72. The distance items were: 'Do you feel negatively about me?', 'Do I leave you with cold feelings?', and 'Do you think it is annoying to deal with me?'. The scale was reliable with a Cronbach's alpha of .81.

Use Intention. Use intention, based on Van Vugt et al. (2006), was measured using 2 indicative items ('Do you want to see me more often on the Internet?'; 'Do you want more information from me in the future?') and 3 counter-indicative items ('Do you want to get rid of me'; 'Would you like me to disappear from the screen?'; 'Would you rather remove me from the screen?'), Cronbach's alpha = .88.

User perceptions. We used a *perceived similarity* scale to check the similarity manipulations. Tversky (1977) showed that similarity is psychologically asymmetrical. That is, if the embodied agent is used as referent (I look like *René*), similarity ratings may be different than when the participant is used as referent (René looks like *me*). Therefore, our perceived similarity scale used items with different referents[4]. To avoid directing the participant in an affirmative answering mode (Dillman, 2000), half of the similarity items were indicative and the other half counter-indicative (reverse-coded). The scale was reliable with a Cronbach's alpha of .93. Further, we used a *perceived idealness* scale to check the ideal similarity manipulations. We used 2 items to measure the extent to which René(e) looked like the participant's ideal or not (perceived idealness), ('Do you want to look like me?'; 'Do you want to look different from me?'). These idealness items correlated sufficiently and significantly (r = .68). Last, *perceived ethics* was measured using 2 items, concerning trustworthiness and credibility (r = .62). Finally, participants answered several questions about personal information: the participant's gender, age, weight, height, general computer experience, ethnicity, and education level.

[4] In the first set of items, the participant was the referent (e.g., 'Do you think I am like you?'). In the second set, the embodied agent was the referent (e.g., 'Do you think you are like me?'). In the third set, there was no explicit referent (e.g., 'Do you think we resemble each other?').

5.2.2 Results

Outlier analysis. Because four participants had outliers on five or more items of various scales, these participants were regarded as unreliable and were disregarded from subsequent analyses. Incidental outliers on individual items were replaced by the mean of the remaining values.

Manipulation checks. We assessed the effectiveness of our manipulations of similarity (similar versus dissimilar body size) and idealness (ideal versus non-ideal body size) by performing a MANOVA with perceived similarity and perceived idealness as dependents. The tests of between-subject effects revealed a significant effect of the similarity conditions on similarity perceptions in the right direction ($F(1, 72) = 14.15$; $p < .001$, partial $\eta^2 = .16$; similar to actual body size $M = 3.1$, $SD = .87$; dissimilar to actual body size $M = 2.2$, $SD = .96$). Furthermore, there was a significant effect of the idealness conditions on perceptions of idealness into the right direction ($F(1, 72) = 18.79$, $p < .001$, partial $\eta^2 = .21$; ideal body size $M = 3.2$, $SD = .89$; non-ideal body size $M = 2.2$, $SD = 1.0$). Thus, we successfully manipulated similarity and idealness by means of varying the body size of an embodied agent.

Testing hypotheses. To test hypothesis H1 versus hypothesis H2, we conducted a MANOVA with similarity (similar versus dissimilar) and idealness (ideal versus non-ideal) as the between-subject factors, and involvement, distance, and use intentions as the dependent variables. Contrary to our expectations, the multivariate test revealed no main effect of similarity ($F(3, 69) = 1.00$; $p = .40$, partial $\eta^2 = .042$), no main effect of idealness ($F(3, 69) = 1.68$; $p = .18$, partial $\eta^2 = .07$), and no interaction effect ($F(3, 69) = 1.15$; $p = .334$, partial $\eta^2 = .05$). Thus, neither similarity nor idealness could significantly predict involvement, distance, and use intentions, indicating no support for hypothesis H1 nor for hypothesis H2. A similar e-health advisor evoked as much use intentions, involvement, and distance as a dissimilar e-health advisor. In addition, a non-ideal e-health advisor evoked as much use intentions, involvement, and distance as an ideal e-health advisor.

Inspecting the research questions. In RQ1, we asked whether the body size of an e-health advisor would affect perceptions of ethics. A correlation analysis showed that perceived ethics correlated significantly with the body size of the e-health advisor that participants interacted with ($r = .24$; $p < .03$). Therefore, we categorized the e-health advisor into three body size groups (according to the Figure Rating Scale sizes 1-3, 4-6, 7-9, see Figure 5.1), and plotted these against the perceptions of ethics in these categories (see Figure 5.3). Thus, the heavier the e-health advisor, the more ethical (here, trustworthy and credible) participants perceived the e-health advisors to be.

In RQ2, we asked whether comparison processes or perceptions of ethics were more important for understanding user responses to e-health advisors. Therefore, we analyzed the data by means of multiple hierarchical regression analyses (HRA's), that is, one HRA for each dependent variable (involvement, distance, use intentions), see Table 5.2. In all three HRA's, the manipulated factors were entered

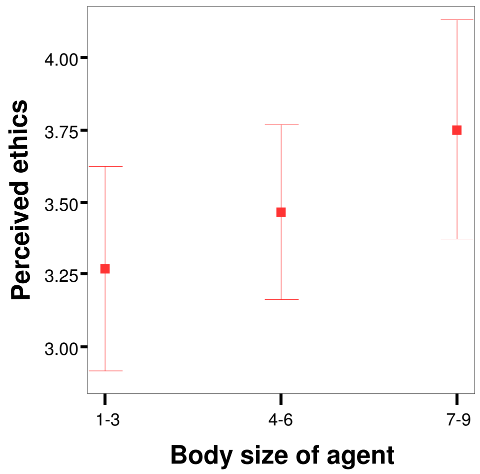


Figure 5.3: Correlation between the body size of the e-health advisor (numbers refer to the Figure Rating Scale), and perceived ethics in Study 1

in Step 1 (i.e., similar versus dissimilar agent, ideal versus non-ideal agent), and the perception of ethics was entered in Step 2.

These analyses confirmed that similarity and ideal similarity did not affect the various user responses, but that *perceived ethics* did. More specifically, the results after Step 2 of each HRA showed that perceived ethics was a significant predictor of involvement, distance, and use intentions (see Table 5.2). That is, perceived ethics accounted for 26% of the variance in involvement, for 11% of the variance in distance and for 21% of the variance in use intention (using the test statistic R^2). Thus, perceived ethics was a better and significant predictor of the various user responses to e-health advisors than the manipulated factors similarity and idealness.

5.2.3 Conclusion and discussion Study 1

The goal of Study 1 was to investigate the effects of similar or ideal body size (as manipulated factors) on user involvement with, distance towards, and intentions to use e-health advisors. Furthermore, we wondered whether the body size of an e-health advisor affected perceptions of ethics (slim being more or less trustworthy), which may subsequently affect involvement, distance and use intentions. Last, we wondered whether comparison processes or perceptions of ethics would be the better predictors of the various user responses to e-health advisors.

Despite the successful manipulations, no direct effects of similarity and idealness were found. Further analyses showed that the body size of the e-health advisor correlated with perceptions of ethics. It appeared that heavier advisors were perceived as more trustworthy and credible than the slimmer ones. The perception of

Table 5.2: Results of Hierarchical Regression Analyses (Step 2) of similarity (similar versus dissimilar), idealness (ideal versus non-ideal), and perceived ethics on involvement, distance, and use intentions, Study 1.

Dependent variable	Variable	*t*	*p*	*Beta*
Involvement	Step 2			
	Similarity (Factor)	.83	.41	.08
	Idealness (Factor)	.31	.76	.03
	Perceived ethics	5.29	.000	**.53**
Note: Step 1 R^2 = .02; Step 2 R^2 = .28				
Distance	Step 2			
	Similarity (Factor)	-.12	.90	-.01
	Idealness (Factor)	1.46	.15	.16
	Perceived ethics	-3.20	.002	**-.34**
Note: Step 1 R^2 = .06; Step 2 R^2 = .17				
Use Intention	Step 2			
	Similarity (Factor)	1.34	.18	.13
	Idealness (Factor)	-.95	.34	-.10
	Perceived ethics	4.724	.000	**.47**
Note: Step 1 R^2 = .07; Step 2 R^2 = .28				

ethics in turn affected the various user responses (involvement, distance, and use intentions) to e-health advisors. Thus, the perception of ethics was indeed dependent on the body size of the e-health advisor, and appeared predict user responses to e-health advisors better than the manipulated factors.

The absence of direct effects of similarity and idealness on user involvement, distance, and use intentions is surprising because well-known and well-established identification theories, such as social comparison theory (e.g., Festinger, 1954; Wills, 1981; Tesser, 1988), state that similarity and idealness are important predictors of identification with others. One reason for the null effects in the present study could be that most participants had thin to normal-sized bodies, only a few were overweight, and none were obese. Regarding body size, the sample was not representative of Dutch society (counting approximately 40% overweight, Visscher et al., 2002), and the actual body sizes of the participants in the similar and ideal condition and those in the similar and non-ideal condition did not differ as much as we intended. This resulted in embodied agents with comparable body sizes among those conditions (*M* = 3.3 and *SD* = .92, versus *M* = 4.2 and *SD* = 1.4, see also Table 5.1). Thus, to create conditions that more clearly differentiated body size, Study 2 included a broader range of participants.

Another explanation for the body shape effects could be a confound with age - bigger bodies may have been perceived as being older, and therefore more trustworthy. In Study 2, we addressed this alternative explanation. Furthermore, we improved some details in the methodology.

5.3 Study 2

In Study 2, we placed a version of the René(e) software used in Study 1 on 'Kennislink'[5], a popular Web site for science communication. We invited general visitors of Kennislink to participate in our study through an announcement and a link to our study on this website. Because Kennislink aims at providing information to people from all classes and backgrounds, we expected to obtain a more representative sample than the student sample in Study 1. About 40 percent of the adult population in the Netherlands is overweight, and 10 percent obese (Visscher et al., 2002). Hence, we expected to have more overweight participants. Indeed, 39% percent of the participants were overweight according to the classification of the World Health Organization (Body Mass Index over 25, see section 5.3.1).

In Study 2, we retested hypothesis H1 versus hypothesis H2. In addition, based on the results of Study 1, we expected that e-health advisors with a larger body size would be perceived as ethically better (more trustworthy and credible) than e-health advisors with a slimmer body size.

5.3.1 Method

Participants and design

A sample of 278 Dutch people completed the study. Adolescents (participants with an age under 18) were excluded from the analyses to reduce effects of maturity, which left us with 259 participants (148 males, 111 females). Of those participants, 39% were overweight (27%) or obese (12%), see Table 5.3 for the BMI ratings of all adult participants.

Table 5.3: The Body Mass Index (BMI) ratings of the adult participants in Study 2.

BMI type	*N*	%
$BMI < 20$ **(underweight)**	24	9%
$20 <= BMI <= 25$ **(normal)**	135	52%
$25 <= BMI <= 30$ **(overweight)**	69	27%
$BMI > 30$ **(obese)**	31	12%
total	259	100%

The same design was used as in Study 1. Assignment of participants to experimental conditions was slightly unbalanced (see Table 5.4), because the 'similar *and* ideal' condition was created after data-collection, and because adolescents were excluded from the data analyses. This sample was more representative of the population than Study 1, given the 40% overweight. The actual body sizes of participants in the similar and ideal condition were significantly smaller than those in the similar and non-ideal condition (BMI ideal M = 25.7, SD = 4.1; BMI non-ideal M = 22.4, SD = 2.7; $t(124) = -4.98$, $p < .001$). As a consequence, the range of

[5]http://www.kennislink.nl

body sizes of the e-health advisors was larger among these conditions (*M* = 3.6, *SD* = .85 versus *M* = 5.1, *SD* = 1.22, see also Table 5.4) than in Study 1. Hence, we expected that effects of the manipulated factors similarity and idealness would be much clearer than in Study 1.

Table 5.4: Number of participants of Study 2 (*n*), means (*M*) and standard deviations (*SD*) of the body size of the used e-health advisors in the similarity and idealness conditions, based on 1-9 range of the Figure Rating Scale.

	Similar to actual self	**Dissimilar to actual self**
Ideal	similar and ideal $n = 51$ M(size) = 3.6, SD = .85	dissimilar but ideal $n = 70$ M(size) = 3.5, SD = .88
Non-ideal	similar but non-ideal $n = 75$ M(size) = 5.1, SD = 1.22	dissimilar and non-ideal $n = 63$ M(size) = 8.2, SD = 1.93

Materials

The materials were almost identical to those used in Study 1. To increase the participants' feeling that the communication with René(e) was actually interactive, they were informed that extended communication with René(e) depended on their input. More specifically, after asking the health-related questions, René(e) told the participants the following: 'I will give you six interesting facts about health. I will ask you every time how much you already know about that fact - quite a lot, a bit, or very little. Depending on how much you know, I tell you some more about it!' An example of such a fact was 'It has been proven that the consumption of 10 grams of pure chocolate a day is good for the heart'. After participants indicated their degree of knowledge on this fact (quite a lot, a bit, or very little), René(e) would always reply the following: 'That the consumption of 10 grams of pure chocolate a day is good for the heart, is probably due to the so called flavenoids contained in chocolate. Flavenoids also exist in for example onion, apple and tea.' Thus, although participants were led to believe that their own knowledge influenced René(e)'s reply, and hence that the application was interactive, in fact it was not.

Procedure

In October and November of 2006, the homepage of the Kennislink Web site contained a link to the René(e) Web site. If participants clicked on the link, a window opened containing our study. As in Study 1, participants were welcomed, asked to enter their gender, weight, height, and age, and to indicate their actual-self, ideal-self, and non-ideal self on three Figure Rating Scales. Then, participants interacted with the embodied agent René(e) for about 8-10 minutes. After the interaction, they completed a user perception questionnaire. Last, they were thanked for participating in the study. They could leave their email addresses (stored separately from the database) if they had questions or wanted more information about the study.

Measurements

As in Study 1, all measurements were taken by means of a questionnaire containing Likert-type scales (1-6).

Dependent measures. Involvement, Distance, and Use Intentions were measured using 4 items each, almost identical to the items used in Study 1. The involvement scale was reliable with a Cronbach's alpha of .86. The distance scale was reliable with a Cronbach's alpha of .80. The use intention scale was reliable with a Cronbach's alpha of .88.

User perceptions. We checked for the similarity manipulations by means of a *perceived similarity* scale. We reduced the set of similarity items of the first study to 2 indicative and 2 counter-indicative items. The scale was reliable with a Cronbach's alpha of .81. We checked for the ideal similarity manipulations by means of a *perceived idealness* scale, using 1 indicative and 1 counter-indicative item, which correlated significantly (r = .80). In addition, the *perceived ethics* scale used in Study 1 was improved and extended to include 4 items, and was reliable (Cronbach's alpha = .67).

As in Study 1, the questionnaire included several items asking for the participants' personal information: gender, age, weight, height, general computer experience, ethnicity, and education level.

Finally, we asked participants to estimate René(e)'s age. In Study 1, heavier agents may have been perceived as older than the slimmer agents, which might have confounded the effect of body size on perceptions of ethics. In other words, an alternative explanation for the correlation between idealness and perceptions of ethics found in Study 1 could be that participants perceived the non-ideal, heavier, e-health advisors as *older* than the thinner ones, and therefore more likely to adhere to moral standards. Therefore, we asked participants to estimate the embodied agent's age, to test whether estimated age correlated with the body size of the agent, and whether estimated age varied among similarity or idealness conditions.

5.3.2 Results

Manipulation checks. We assessed the effectiveness of our manipulations of similarity (similar versus dissimilar body size) and idealness (ideal versus non-ideal body size) by performing a MANOVA with perceived similarity and perceived idealness as dependents. The tests of between-subject effects revealed a significant effect of the similarity conditions on similarity perceptions in accordance with our intentions ($F(1, 255) = 71.06$; $p < .001$, partial $\eta^2 = .22$; similar to actual body size $M = 3.6$, $SD = .95$; dissimilar to actual body size $M = 2.6$, $SD = .90$). Furthermore, there was a significant effect of the idealness conditions on idealness perceptions in the right direction ($F(1, 255) = 111.29$, $p < .001$, partial $\eta^2 = .30$; ideal body size

M = 3.9, SD = 1.0; non-ideal body size M = 2.6, SD = 1.1). Thus, we successfully manipulated similarity and ideal similarity by means of varying the body size of an embodied agent.

Body size and age estimates. We tested whether the estimated *age* of René(e) was dependent on body size. We computed the correlation coefficient between the body size of the agent that participants interacted with and estimated age. The agent's body size did not significantly correlate with estimated age (r = .50; p = .44). Then, to verify that the estimated age could not explain effects among conditions, we also tested whether the estimated age varied among the conditions. Therefore, we performed an ANOVA with similarity (similar versus dissimilar) and idealness (ideal versus non-ideal) as the between-subject factors, and estimated age as the dependent variable. There was no main effect of similarity ($F(1, 255) = 1.51$, p = .22, partial η^2 = .006), no main effect of idealness ($F(1, 255) = 1.25$, p = .27, partial η^2 = .005), and no interaction effect ($F(1, 255) = .044$, p = .83, partial η^2 = .000). More specifically, the ideal and non-ideal, and the similar and dissimilar agents were all estimated as being about 30 to 40 years of age. Thus, estimations of René(e)'s age did not correlate with René(e)'s body size, nor did it differ among conditions, and could therefore not explain the effects of the manipulations.

Testing hypotheses[6]. As in Study 1, to test hypothesis H1 versus hypothesis H2, we performed a MANOVA with similarity (similar versus dissimilar) and idealness (ideal versus non-ideal) as the between-subject factors. The dependent variables were use intentions, involvement, and distance. The multivariate test showed a marginal effect of the factor similarity ($F(3, 253) = 2.48$, p = .061, partial η^2 = .03), a main effect of the factor idealness ($F(3, 253) = 3.17$, $p < .025$, partial η^2 = .04), and no interaction effect ($F(3, 253) = .234$, p = .87, partial η^2 = .003). The univariate tests revealed that the marginal multivariate effect of similarity remained insignificant for all dependents (ps $> .20$). Hence, similarity did not affect the dependents. The univariate tests further revealed that idealness affected intentions to use the e-health advisor ($F(1, 255) = 4.06$, $p < .045$, partial η^2 = .16), as well as distance towards the e-health advisor ($F(1, 255) = 4.86$, $p < .028$, partial η^2 = .19), but not involvement with the e-health advisor ($F(1, 255) = .28$, p = .60, partial η^2 = .001). Participants were more eager to use and felt less distant towards the non-ideal e-health advisors than towards the ideal e-health advisors. Thus, we found support for hypothesis H2, but not for hypothesis H1. These results indicated that idealness was more important than similarity in explaining user responses to e-health advisors.

Inspecting the research questions. As in Study 1, to inspect RQ1, we tested whether the body size of the e-health advisor and perceptions of ethics were correlated. A correlation analysis showed that perceived ethics correlated significantly with the body size of the e-health advisor that participants interacted with (r = .16; $p < .01$), but not very strongly. Consistent with findings in Study 1, the heavier the e-health advisor, the more ethical participants perceived the e-health advisors to be

[6]There were no outliers in the data.

(i.e., the more trustworthy and credible).

Then, as in Study 1, to inspect RQ2, we analyzed the data by means of multiple hierarchical regression analyses (HRA's), that is, one HRA for each dependent variable (involvement, distance, use intentions), see Table 5.5. In all three HRA's, the manipulated factors were entered in Step 1 (i.e., similar versus dissimilar agent, ideal versus non-ideal agent) and the perception of ethics was entered in Step 2.

Table 5.5: Results of Hierarchical Regression Analyses (Step 2) of similarity (similar versus dissimilar), idealness (ideal versus non-ideal), and perceived ethics on involvement, distance, and use intentions, Study 2.

Dependent variable	Variable	*t*	*p*	*Beta*
Involvement	Step 2			
	Similarity (Factor)	-1.39	.17	-.072
	Idealness (Factor)	-.07	.51	-.04
	Perceived ethics	11.14	.000	**.58**
Note: Step 1 R^2 = .01; Step 2 R^2 = .33				
Distance	Step 2			
	Similarity (Factor)	-1.39	.17	-.08
	Idealness (Factor)	-1.49	.14	-.08
	Perceived ethics	-8.90	.000	**-.48**
Note: Step 1 R^2 = .15; Step 2 R^2 = .25				
Use Intention	Step 2			
	Similarity (Factor)	-.31	.75	-.02
	Idealness (Factor)	1.18	.24	.06
	Perceived ethics	10.14	.000	**.54**
Note: Step 1 R^2 = .02; Step 2 R^2 = .30				

The results after Step 1 of each HRA were consistent with the above MANOVA results. Idealness affected both distance and use intentions, but not involvement, and similarity did not affect any of the dependents. The results after Step 2 of each HRA showed that the perception of ethics was a significant predictor of involvement, distance, and use intentions (see Table 5.5). That is, the perception of ethics accounted for 32% of the variance in involvement, for 10% of the variance in distance and for 28% of the variance in use intention (using the test statistic R^2). However, effects of the manipulated factors similarity and idealness became non-significant. Thus, the perception of ethics overruled the effect of idealness. As in Study 1, the perception of ethics was a better and significant predictor of the various user responses to e-health advisors than the manipulated factors similarity and idealness. Hence, perceived ethics was more important than similarity and idealness were for user responses to e-health advisors.

The results after Step 2 of each HRA showed that the perception of ethics was a significant predictor of involvement, distance, and use intentions (see Table 5.5). That is, the perception of ethics accounted for 32% of the variance in involvement, for 10% of the variance in distance and for 28% of the variance in use intention (using the test statistic R^2). Effects of the manipulated factors similarity and idealness became non-significant. Thus, the perception of ethics overruled the effect of idealness. As in Study 1, the perception of ethics was a better and significant predictor of the various user responses to e-health advisors than the manipulated factors similarity and idealness. Hence, perceived ethics was more important than similarity and idealness were for user responses to e-health advisors.

5.3.3 Conclusion Study 2

The goal of Study 2 was the same as that in Study 1. However, Study 2 included a more representative sample of Dutch society (with approximately 40% of people overweight). We used an *online* setting, and 39% of participants were overweight. Study 2 found a direct effect of idealness on intentions to use and on the reduction of distance. The non-ideal (heavier) e-health advisors relieved emotional distance and boosted usage more than ideal, slim e-health advisors did. However, no effects were found for similarity. Thus, (non)idealness was more important than similarity for explaining user distance and use intentions. Whereas hypothesis H2 (concerning idealness) was supported by the data, hypothesis H1 (concerning similarity) was not supported. In addition, we found that body size significantly correlated with perceptions of ethics - the heavier the advisor, the more trustworthy and credible he or she was perceived to be. Last, perceived ethics was a better predictor of the various user responses than idealness. Body size and age estimation were uncorrelated. Thus, estimated age could not have confounded the results.

In Study 2, we found direct effects of idealness on user responses, which were not found in Study 1. A plausible reason for this difference is that in Study 2, more people participated in the experiment, and they had greater variation in body sizes than the participants in Study 1, resulting in both larger and more extreme experimental conditions. Non-ideal (heavier) e-health advisors evoked less negative and more positive user responses than ideal (thinner) e-health advisors. Hence, we concluded that heavier embodied agents are better e-health advisors than thin ones.

5.4 General conclusion and discussion

The reported studies indicate that people compare e-health advisors not only with their *actual* selves, with their *ideal* selves even more. One laboratory and one field study scrutinized the effects of similarity and idealness of an embodied agent feature on user involvement with, distance towards, and intentions to use the agent. The agent's body size was either similar or dissimilar to the user's actual body size, and the body size was either ideal or non-ideal. In both studies, participants interacted with an embodied agent in an e-health context. We found idealness to be more important than similarity for explaining user responses. Although people may feel attracted to the similarity they perceive in embodied agents, this effect might be overruled by the perception of idealness (cf. Wetzel and Insko, 1982; LaPrelle et al., 1990). An embodied agent may possess an ideal feature that users look up to (evoking feelings of approach), or that elicits frustration, for example out of jealousy (evoking feelings of avoidance). This is particularly interesting because similarity has been investigated in a range of embodied agent studies, whereas idealness has been mostly neglected. The effects that idealness may have on the user deserves more attention in embodied agent research.

Ethical considerations appeared to be even more important than idealness. In both studies, the heavier the e-health advisors were, the more moral they were perceived to be. As a result, the heavier e-health advisors were more involving, less distancing, and evoked higher use intentions than slimmer e-health advisors. Whereas stereotype theory predicts that 'What is beautiful is good' (Dion et al., 1972), the present study showed that this stereotype is not prevalent in all contexts. That is, in a context related to body size, such as a health context where food intake and diets are relevant topics, an ideally thin (considered beautiful) body size may lead to counter-intuitive effects. In other words, the effects of embodied agents on users might not be what general common-sense reasoning and prevailing stereotypes would lead us to expect. A non-ideal, heavier body shape may preferable in e-health contexts. Thus, sometimes it is better to design embodied agents that are not ideal.

Apparently, processes of comparison that are known to occur in face-to-face communication, also exist when the 'other' is merely a software agent. This is consistent with the well-known media equation studies, which demonstrate that users treat computers like real people (Reeves and Nass, 1996). Based on downward comparison theory (e.g., Wills, 1981; Suls et al., 2002), ideally thin users probably felt better interacting with a non-ideal e-health advisor than with an ideal e-health advisor because the heavier agents were 'worse-off' than themselves. Apparently, when comparison concerns an important feature for self-definition, it feels good to look down on others, even if they are software agents. Based on upward comparison theory (e.g., Bers and Rodin, 1984; Tesser, 1986, 1988; Salovey and Rodin, 1991), heavier users probably turned away from an e-health advisor with an ideal body size because it threatened self-evaluation and evoked jealousy. In other words, a theoretical explanation might be that heavier people dissociate from the ideal agent to maintain high self-esteem about their own appearance.

Another reason that heavier users prefer heavier agents could be that the doctor's body size may influence the *magnitude* of his or her anti-fat bias. Research has shown that, although all weight groups exhibit an anti-fat bias (there is no *pro*-fat bias among obese people), the magnitude of the anti-fat bias is weaker among people with a high Body Mass Index (BMI) compared to a low Body Mass Index (Schwartz et al., 2006). Thus, heavier people adopt the negative stereotype towards heavy people to a lesser extent than thin people do. In a similar vein, a heavy health advisor may adopt the negative stereotype towards heavy people to a lesser extent than thin advisors. Thus, heavy people may expect that heavier health advisors understand their situation better than thinner health advisors: 'The doctor himself is overweight, so s/he will understand how I feel!' In addition, they may be more likely to believe an advisor that points out the negatives of overweight when the advisor is heavy him- or herself.

The method we used appeared effective to study both similarity and idealness effects on the same dimension (body shape), at the same time. We explicitly asked the participants in the studies to indicate their actual body size, their ideal body size and their non-ideal body size. Participants did not confuse actual with ideal body sizes, because actual body size is a fact. Sometimes, however, it is difficult to separate the effects of similarity from those of (non-)ideal similarity (Herbst et al., 2003). For example, it is difficult to distinguish between the effects of similarity to the actual attitude and similarity to the ideal attitude because people can change their attitudes towards the ideal. Such confusion is less likely to happen for traits, such as extraversion, or for physical features, such as body size, that are more difficult to change (cf. Herbst et al., 2003).

5.4.1 Limitations and future studies

The null effects of similarity might have occurred for a variety of reasons. One reason might be that participants in our samples showed little variance in perceived similarity. Perceived similarity ratings showed that participants did not feel very similar to the embodied agents in any of the conditions, although perceptions of similarity differed significantly between the conditions. However, similarity effects did occur in one of our previous studies in which participants' perceptions of similarity were equally low (Van Vugt et al., 2007a). Therefore, a more plausible reason for the null effect is that effects of similarity were *overruled* by effects of ideal similarity. Idealness was the stronger predictor, which has also occurred in social psychology studies (e.g., Herbst et al., 2003). It would be worthwhile to study not only actual similarity, but also ideal similarity in future work on embodied agents. Researchers may study similar versus ideal personalities, similar versus ideal attitudes, similar versus ideal (nonverbal) behaviors, and similar versus ideal bodily features, such as facial features or costumes.

In general, our findings indicate that designers should be cautious using 'ideal' agents by default, especially when feature(s) are considered self-relevant by users in a certain context, such as the body size feature in a health context. Other agent features such as personality, attitudes, behavior, or other bodily features, may be considered self-relevant in other contexts, such as personality and looks on an online dating site. It remains to be investigated whether users prefer an agent that is non-ideal on these features in certain contexts. However, only few features of an agent are likely to be relevant in a certain context of use, and for the irrelevant features, designing 'ideal' agents is presumably the way to go. For example, in applications that do not explicitly deal with body size, it might be better to design embodied agents with an ideal, slim, body size. Virtual news readers, virtual storytellers, and virtual sales assistants (that work in a context in which body size is not relevant) may be more effective with a thin body.

Furthermore, perceptions of what is ideal and non-ideal change through times and places. For example, being oversized is a sign of well-being and something to strive for in Morocco and other African countries (Rguibia and Belahsen, 2006). Cultural differences between what is perceived as ideal or non-ideal also occur on other dimensions such as personality, attitudes, and (non-verbal) behaviors (e.g., Tsai et al., 2007; Singh et al., 1998). Thus, users of different cultural backgrounds are likely to respond differently to embodied agents with certain appearances, personalities, attitudes, and behaviors. If African people see an e-health advisor with a slim body, they might perceive it as non-ideal (worse-off) rather than ideal (better-off), shaping their responses accordingly. What is ideal for one group may be non-ideal for another. This is especially important when designing for a multi-cultural public.

In the present study, perceptions of ethics were affected by the body size of the e-health advisor. However, whether an embodied agent (e.g., with a different role) is perceived as morally good may not only depend on its body size. The ethnicity and age of an agent, the reputation of the company an agent is working for, the moral behavior of an agent (e.g., spyware, spam behavior), an agent's verbal and non-verbal interaction style, and the personalization of content may also affect perceptions of ethics (e.g., Cowell and Stanney, 2005; Sillence et al., 2006). The formation of perceptions of the ethical intentions of agents might be an interesting topic for future research, especially because ethical considerations go beyond health systems, and are also important for users of, for example, e-commerce, advice, and recommender systems (e.g., Wang and Emurian, 2005).

Finally, the present studies examined *intentions* to use e-health advisors but did not assess participants' actual efforts to use an agent again. Although intentions to use and actual use are highly related (Davis, 1989), future research should also include measures of actual use. Advice following, the use of prescribed medication, patient health outcomes, and patients' reported satisfaction with the medical care could be examined as additional dependent measures in further studies. The results of future studies may reveal more unexpected implications for the design of trustworthy embodied agent systems, just like the present study did.

5.5 Acknowledgments

We are grateful to the Faculty of Social Sciences and the Faculty of Sciences of the VU University in Amsterdam, who granted a 4-year research track (VUBIS) to the first author. Further, we thank Kennislink for posting a link to our study on their Web site. This enabled us to perform the second study. We also thank Erik Kruithof for implementing the René(e) software and for his help during the experiments. Last, we thank Jan Kleinnijenhuis and Gerrit van der Veer for their support.

Bibliography

Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., and Weisbuch, M. (2001). Intelligents agents who wear your face: User's reactions to the virtual self. In A. de Antonio, R. Aylett, D. B., editor, *Intelligent virtual agents*, volume 2190 of *Lecture Notes in Artificial Intelligence*, pages 86–99. Springer.

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology*, 3(3):265–299.

Bers, S. A. and Rodin, J. (1984). Social-comparison jealousy: A developmental and motivational study. *Journal of Personality and Social Psychology*, 47(4):766–779.

Bickmore, T., Gruber, A., and Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educational Counseling*, 59(1):21–30.

Bickmore, T. W. and Cassell, J. (2001). Relational agents: A model and implementation of building user trust. In *CHI 2001 Conference Proceedings*, pages 396–403.

Bulik, C. M., Wade, T. D., Heath, A. C., Martin, N. G., Stunkard, A. J., and Eaves, L. J. (2001). Relating body mass index to figural stimuli: population-based normative data for caucasians. *International Journal of Obesity*, 25(10):1517–1524.

Calnan, M. and Rowe, R. (2006). Researching trust relations in health care: Conceptual and methodological challenges? An introduction. *Journal of Health, Organisation and Management*, 20(5):349–358.

Cowell, A. J. and Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies. Special issue: Subtle expressivity for characters and robots*, 62(2):281–306.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–339.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. John Wiley and Sons, New York, 2th edition.

DiMatteo, M. R. and Hays, R. (1980). The significance of patients' perceptions of physician conduct: a study of patient satisfaction in a family practice center. *Journal of Community Health*, 6(1):18–34.

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3):285–290.

Dryer, D. C. (1999). Getting personal with computers: How to design personalities for agents. *Applied Artificial Intelligence*, 13(3):273–295.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7:117–140.

Fingeret, M. C., Gleaves, D. H., and Pearson, C. A. (2004). On the methodology of body image assessment: The use of figural rating scales to evaluate body dissatisfaction and the ideal body standards of women. *Body Image*, 1:207–212.

Greenberg, B. S., Eastin, M., Hofshire, L., Lachlan, K., and Brownell, K. D. (2003). Portrayels of overweight and obese individuals on commercial television. *American Journal of Public Health*, 93(8):1342–1348.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., and McCall, C. (2007). Virtual humans and persuasion: the effects of agency and behavioural realism. *Media Psychology*, 10:1–22.

Harrison, K., Taylor, L. D., and Marske, A. L. (2006). Women's and men's eating behavior following exposure to ideal-body images and text. *Communication Research*, 33(6):507–529.

Herbst, K. C., Gaertner, L., and Insko, C. A. (2003). My head says yes but my heart says no: cognitive and affective attraction as a function of similarity to the ideal self. *Journal of Personality and Social Psychology*, 84(6):1206–1219.

Hoffner, C. and Buchanan, M. (2005). Young adults wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, 7:325–351.

Kanzler, M. H. and Gorsulowsky, D. C. (2002). Patients' attitudes regarding physical characteristics of medical care providers in dermatologic practices. *Archives of Dermatology*, 138(4):463–466.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A., Nije Bijvank, M., and Bushman, B. (2007). I wish I were a warrior: the role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, 43(4):1038–1044.

LaPrelle, J., Hoyle, R. H., Insko, C. A., and Bernthal, P. (1990). Interpersonal attraction and descriptions of the traits of others: Ideal similarity, self similarity, and liking. *Journal of Research in Personalit*, 24:216240.

Lenhart, A., Rainie, L., and Lewis, O. (2001). Teenage life online. The rise of the instant-message generation and the internet's impact on friendships and family relationships. Retrieved May 2007 from http://www.pewinternet.org.

Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103.

Nowak, K. L. and Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1):article 8.

Parker, D. (2007). Copyright: Original Artist. Catalogue Ref: dpan1436. Retrieved September 24, 2007 from http://www.cartoonstock.com/ newscartoons/ cartoonists/ dpa/ lowres/ dpan1436l.jpg.

Piette, J. D., Heisler, M., Krein, S., and Kerr, E. A. (2005). The role of patient-physician trust in moderating medication nonadherence due to cost pressures. *Archives of Internal Medicine*, 165:1749–1755.

Puhl, R. and Brownell, K. (2001). Bias, discrimination and obesity. *Obesity Research*, 9:788805.

Reeves, B. and Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge University Press.

Rehman, S. U., Nietert, P. J., Cope, D. W., and Kilpatrick, A. O. (2005). What to wear today? Effect of doctor's attire on the trust and confidence of patients. *The American Journal of Medicine*, 118(11):1279–1286.

Rguibia, M. and Belahsen, R. (2006). Body size preferences and sociocultural influences on attitudes towards obesity among Moroccan Sahraoui women. *Body Image: An International Journal of Research*, 3(4):395–400.

Salovey, P. and Rodin, J. (1991). Provoking jealousy and envy: Domain relevance and self-esteem threat. *Journal of Social and Clinical Psychology*, 10:395–413.

Schwartz, M., Vartanian, L., Nosek, B., and Brownell, K. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity*, 14(3):440–447.

Shapiro, M. A. and McDonald, D. G. (1992). I'm not a real doctor, but I play one in virtual reality: Implications of virtual reality for judgments about reality. *Journal of Communication*, 42:94–114.

Sillence, E., Briggs, P., Harris, P., and Fishwick, L. (2006). A framework for understanding trust factors in web-based health advice. *International Journal of Human-Computer Studies*, 64:697–713.

Singh, N. N., McKay, J. D., and Singh, A. N. (1998). Culture and mental health: Nonverbal communication. *Journal of Child and Family Studies*, 7(4):403–409.

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: a review. *Canadian Medical Association Journal*, 152(9):1423–1433.

Stunkard, A. J., Sorenson, T., and Schlusinger, F. (1983). Use of the danish adoption register for the study of obesity and thinness. In S.S. Kety, L.P. Rowland, R. S. and Matthysse, S., editors, *The genetics of neurological and psychiatric disorders*, pages 115–120, New York. Raven Press.

Suls, J., Martin, R., and Wheeler, L. (2002). Social comparison: Why, with whom and with what effect? *Current Directions in Psychological Science*, 11(5):159–163.

Teachman, B. A. and Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: is anyone immune? *International Journal of Obesity*, 25:1525–1531.

Tesser, A. (1986). Some effects of self-evaluation maintenance on cognition and action. In Sorrentino, R. M. and Higgins, E. T., editors, *Handbook of motivation and cognition: Foundations of social behavior*, pages 435–464, New York. Guilford.

Tesser, A. (1988). Towards a self-evaluation maintenance model of social behavior. In Berkowitz, L., editor, *Advances in experimental social psychology*, volume 21, pages 181–227, San Diego. Academic Press.

Tsai, J. L., Louie, J. Y., Chen, E. E., and Uchida, Y. (2007). Learning what feelings to desire: Socialization of ideal affect through children's storybooks. *Personality and Social Psychology Bulletin*, 33(1):17–30.

Tversky (1977). Features of similarity,. *Psycological Review*, 84(4).

Van Vugt, H. C., Bailenson, J. N., Hoorn, J. F., and Konijn, E. A. (2007a). Facial similarity between embodied agents and users: the differential effects of affordances and gender. submitted.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I., and Eliëns, A. (2007b). Realism is not all! User engagement with task-related interface characters. *Interacting with Computers*, 19(2):267–280.

Visscher, T. L., Kromhout, D., and Seidell, J. C. (2002). Long-term and recent time trends in the prevalence of obesity among dutch men and women. *International Journal of Obesity and Related Metabolic Disorders*, 26(9):1218–1224.

Wang, Y. D. and Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, 21:105–125.

Wetzel, C. G. and Insko, C. A. (1982). The similarity-attraction relationship: Is there an ideal one? *Journal of Experimental Social Psychology*, 18(9):253276.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90:245 271.

World Health Organisation (2000). Obesity - preventing and managing the global epidemic. Report of a WHO consultation. Technical report series no. 894, WHO, Geneva, Switzerland.

World Health Organisation (2003). Obesity and overweight. Retrieved June, 19, 2007 from http://www.who.int/hpr/NPH/docs/.

CHAPTER 6

Conclusions and Discussion

The central question this dissertation addressed is how computer users respond to embodied agents. We considered satisfaction level to be an important end-response of user-agent interactions. End-user satisfaction is an important goal in user-system interaction and design. Because it evaluates the overall, 'summative' performance of a system, the measure can indicate which systems are 'better' than others (Nielsen, 1993; Walker et al., 1997; Lindgaard and Dudek, 2003; Sanders and Scholtz, 2000). Designers may assume that user satisfaction depends on design features alone (cf. Han and Hong, 2003). Hence, designers of embodied agents focus on including advanced features, such as models of behavior, emotion, speech, and intelligent reasoning, in addition to realistic appearances (e.g., Dehn and Van Mulken, 2000; Cerrato and Ekeklint, 2004; Sanders and Scholtz, 2000; Gratch and Marsella, 2005; Pelachaud et al., 2007). However, user satisfaction is not shaped solely by design, but more importantly, by the user. Investigating embodied agent systems from the user's perspective is crucial for a deeper understanding of *why* users feel satisfied or dissatisfied with an embodied agent.

First, we developed the Interactive PEFiC model, I-PEFiC, that focuses on the psychological processes embodied agents evoke in users (cf. section 1.5). Hence, the first contribution of this dissertation is a conceptual one: I-PEFiC as a framework for the study of human-agent interactions. The conceptual model was based on the PEFiC model for perceiving and experiencing fictional characters that was empirically validated with film and television characters and media figures (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005; Konijn and Bushman, 2007). Importantly, these studies found that *both* involvement *and* distance experiences, the two dimensions of engagement (Konijn and Hoorn, 2005; Konijn and Bushman, 2007) explain the appreciation of a mediated person. Thus, involvement and distance are distinct experiences that are not two ends of a single dimension; both can be experienced at the same time. The trade-off between involvement and distance better explains appreciation of a fictional character than either involvement or distance alone (Konijn and Hoorn, 2005; Konijn and Bushman, 2007). Therefore, we treated involvement and distance as separate dimensions of engaging with an embodied agent and tested the validity of this distinction.

Because people can interact with embodied agents, we adjusted the PEFiC model to incorporate the interactive nature of human-agent encounters, using theories from HCI. The model assumes that users perceive embodied agents as fictional characters as well as tools, and that encounters with an embodied agent trigger both an engagement process and an interaction process. Both processes occur simultaneously and contribute to the user's final satisfaction. In sum, I-PEFiC aims to explain not only involvement with and distance towards, but also intentions to use, and final satisfaction with embodied agents.

In following the PEFiC model, I-PEFiC also assumes that user perceptions of ethics, aesthetics, and realism typically evoke involvement and distance, and that perceptions of affordances typically evoke intentions to use an embodied agent. Perceptions of similarity, relevance, and valence are expected to mediate between perceptions of ethics, aesthetics, realism and affordances on the one hand, and user responses of involvement, distance, and use intentions on the other.

This theoretical I-PEFiC model was used as a framework for empirical research. Several studies tested important parts of I-PEFiC empirically. An overview of the factors manipulated and measured in the studies may be found in Chapter 1, Table 1.1. Because the research literature suggests that realism is a key factor influencing user responses to embodied agents, in Chapter 2 we focused on realism, residing from the character-side of I-PEFiC. In Chapter 3, we focused on the effects of combinations of a factor from the character-side, aesthetics, with a factor from the system-side, affordances. In Chapter 4, we further studied the effects of combinations of a factor from the character-side, facial similarity, with affordances. In Chapter 5, we continued to study similarity effects, but as it related to body shape.

First, this final chapter summarizes results from the empirical studies. Then, it presents the main conclusions that can be drawn from these studies, and how those conclusions suggest the 'revised I-PEFiC model'. Based on the empirical results, this new model describes, more accurately than the original I-PEFiC model, how users perceive and respond to embodied agents. Revised I-PEFiC may inform agent design decisions and may serve as a starting-point for future research activities concerning embodied agents or related domains such as robotics.

6.1 Summary of the empirical results

I-PEFiC considered satisfaction to be the end-response of user-agent interactions. Indeed, as predicted, both an engagement process (the character-side of user-agent interactions) and an interaction process (the system-side of user-agent interactions) contribute to end-user satisfaction (section 2.4.2 and section 3.4.3. Involvement, distance, and use intentions significantly contributed to satisfaction. To feel satisfied, a user should feel involved with, not too distant from, and have intentions to use the agent. In addition, psychometric analyses performed in Chapter 4 (in section 4.2.7) showed that engagement (involvement and distance, emotional approach or avoidance) is not the same as use intentions (actual approach or avoidance). Feeling involved with an agent does not necessarily imply actual use, and feeling distant from an agent does not necessarily imply non-use (cf. Saffer, 2007, page 15 of this dissertation). The summary below presents the results of the empirical studies concerning engagement (involvement and distance) and use intentions as dependent variables, see also Table 6.1 on page 157.

Chapter 2 presented an initial empirical study in which participants were confronted with an agent, a virtual museum guide of either realistic or unrealistic appearance, or no agent at all. Each participant's task was to learn as much as possible about an art installation. We wondered whether the agent's presence would affect task performance (e.g., 'With a guide, I remember more'), and we expected that the agent's features would affect participants' engagement with the agent (e.g., 'If the guide looks unrealistic, I feel less involved').

We found no effect of agent presence on task performance. Regardless of whether an embodied agent was present, task performance was the same. In addition, although the level of realism of the embodied agent and perceived aesthetics contributed to user engagement, these factors did not affect how much participants remembered from the tour. Yet, perceived relevance of the agent for the task *did* affect user engagement (here, distance). Thus, it seemed that typical factors for the character-side of I-PEFiC (e.g., realism and aesthetics, see Figure 1.3) did not affect the performance- or task-side of the model but that the reverse *did* occur.

Chapter 3 more closely examined the cross-over of effects between the character- and system-side of I-PEFiC, by systematically manipulating both the embodied agent's outer appearance and affordances as a tool for performing a task. Using the game 'Sims 2', we conducted an experiment to test the effects of aesthetics (beautiful vs. ugly) and affordances (help vs. obstacle) on use intentions, user engagement (involvement and distance), and satisfaction.

Results showed that the outer appearance of the embodied agent (here, aesthetics) affected user engagement, but did not affect intentions to use the agent neither directly, nor indirectly via relevance and valence. Moreover, user perceptions of how real the agent's outer appearance seemed contributed to engagement, but not to intentions to use the agent. Thus, as in Chapter 2, we found an expected impact within the character-side, however, cross-over effects to the system-side remained absent. In addition, whether the embodied agent helped or hindered task completion affected not only use intentions but user engagement as well. Participants were least involved with and most distant to the ugly character if it obstructed task completion. If a character aided task completion, however, its helpful affordances compensated for ugliness. Thus, we found an expected impact within the system-side itself, but also a strong cross-over effect to the character-side. Moreover, perceptions of relevance and valence of user goals mediated the relationship between perceived affordances and use intentions, as well as between perceived affordances and engagement (here, distance). All of these findings imply an 'asymmetry effect' - the system- or task-side of I-PEFiC affected the character-side, whereas the reverse did not occur.

Furthermore, an interesting trend (see Figure 3.5) indicated that users were least intended to use an agent when it was unhelpful (understandably) and beautiful (surprisingly)! In addition, when an agent was helpful, participants seemed more willing to use a beautiful than an ugly agent.

Therefore, in Chapter 4 we studied whether a positive feature from the character-side of I-PEFiC (here, facial similarity) could negatively affect intentions to use an obstructing agent. We also intended to study the asymmetry effect further by systematically manipulating the outer appearance of an embodied agent (its facial similarity) as well as the quality of the agent's affordances for use in a task. Unlike the previous studies, the embodied agent existed in a 3D immersive virtual environment (Blascovich et al., 2002).

Results showed that, again, affordances (help vs. obstacle) were more decisive than the outer appearance of an embodied agent (here, facial similarity) for use intentions, involvement, and distance. In general, participants were more eager to use, felt more involved with, and less distant towards an aiding agent compared to an obstructing one. Thus, as in Chapters 2 and 3, the system-side affected the character-side of I-PEFiC. Results furthermore showed that, as in Chapter 3, perceived relevance and perceived valence mediated between perceived affordances and use intentions, as well as between perceived affordances and engagement (distance).

An embodied agent's outer appearance (here, facial similarity) also affected user engagement (here, involvement) with the agent, but to a lesser extent than affordances did. Usually, users 'love their face in virtual space'. However, we found no direct effects of facial similarity on use intentions in our participant sample. Nor did we find indirect effects of facial similarity, via relevance and valence of user goals. Whether facial similarity affected use intentions or not depended on participant gender. Female participants' use intentions were unaffected by facial similarity. Male participants, however, preferred to use a facially similar agent as long as facial similarity was paired with high quality advice (aiding affordances). In addition, male participants preferred to use a facially *dis*similar agent when facial similarity was paired with low quality advice (obstructing affordances). Thus, when a man's agent is an incompetent advisor, it should not look like him. The effects of similarity on involvement (for the whole sample) as well as on use intentions (for males only) were especially remarkable because participants were *unaware* of the similarity manipulation. Thus, in contrast to the previous studies, the character-side of I-PEFiC affected the system-side - but at an unconscious level.

The discussion section 4.4.1 in Chapter 4 suggests several potential explanations for the gender effect that men, more than women, wanted to avoid interacting with an obstructing agent that looked like them. We reasoned that the effect might have been the result of differences between men and women in their achievement needs, in how important they think the usefulness of systems is, and in their task performance expectations. Gender may also be conceptualized as a psychological construct, as opposed to a dichotomous treatment of gender as ‘biological sex’ (e.g., Bem, 1981, 1993; Brosnan, 1998). Thus, gender may be a surrogate for other psychological constructs or dimensions (such as needs and expectations). This is important theoretically, because I-PEFiC is a psychological model. Although treating gender as a biological construct is consistent with previous conceptualizations, it adds a factor to I-PEFiC that might be alleviated by a psychological examination of gender (for a similar reason, the TAM model does not add gender as an additional factor to the model, Venkatesh, 2000). Future research may study whether the encountered gender effects are indeed a substitute for effects of one or several psychological constructs. This would be useful because men and women are not at bipolar extremes on the psychological construct of gender but might vary according to degrees of femininity or masculinity (Bem, 1981, 1993).

In Chapter 5, we performed two studies to investigate the effects of the outer appearance of agents (body shape similarity and idealness) on user responses. Moreover, we further studied whether a positive feature of an agent from the character-side of I-PEFiC (here, body shape similarity and body shape idealness) could have a negative effect on the user, and vice versa. The agent was designed as an e-health advisor, and participants were led to believe that the agent was responding to their input interactively.

Results indicated that idealness was more important than similarity for explaining involvement with, distance towards, and intentions to use the embodied agent, but in unexpected ways. That is, first, the idealness of the agent’s body shape affected engagement and use intentions, but this only was an indirect effect, via perceptions of ethics (is the agent perceived as good or bad, trustworthy or untrustworthy). These results suggest the cross-over of information from the character-side (i.e. idealness, via ethics) to the system-side of I-PEFiC. Second, in contrast to what one might expect from prevailing anti-fat norms in Western society, users perceived the heavier agents (e-health advisors, in this case) as more trustworthy. As a result, the heavier e-health advisors were more involving and less distancing than slimmer e-health advisors, and they evoked higher use intentions as well. Thus, we found that under certain circumstances, a negative feature of an agent may have a positive effect on the user.

Based on the analyses performed in Chapter 5, the roles of relevance and valence in this study are still unclear. Therefore (and to be consistent with Chapters 3 and 4), we conducted post-hoc tests to examine whether perceived relevance and perceived valence were mediators in the studies presented in Chapter 5. The results of three mediation analyses on the data of the more extensive Study 2 in Chapter 5 showed that both perceived relevance and perceived valence mediated the effect of perceived ethics on all three dependents, involvement[1], distance[2], and use intentions[3]. Additional post-hoc mediation analyses using the data of Study 1 in Chapter 5 obtained similar results. Thus, perceived ethics affected involvement, distance, and use intentions indirectly, through relevance and valence. This is a notable finding, because no evidence previously had indicated that relevance and valence mediate the relationship between realism, aesthetics, and facial similarity, and user responses. Thus, ethics, a factor from the character-side of I-PEFiC, plays a more similar role to affordances, a factor from the system-side of I-PEFiC, than to the other factors from the character-side of I-PEFiC, which, unlike ethics, concern the outer appearance of the agent.

The results are summarized in Table 6.1.

[1]relevance: Sobel $z = 8.85$, $p < .001$; valence: Sobel $z = 9.87$, $p < .001$
[2]relevance: Sobel $z = -8.08$, $p < .001$; valence: Sobel $z = -6.96$, $p < .001$
[3]relevance: Sobel $z = 10.61$, $p < .001$; valence: Sobel $z = 9.37$, $p < .001$

Table 6.1: Overview of effects found in the studies

	Manipulations character-side	**User perceptions character-side**	**Effect on character-side?**	**Effect on system-side?**	**Comment**
Chapter 2	(Form) Realism	Ethics, Aesthetics, Realism, Similarity	Yes. Perceived aesthetics had a larger impact on engagement than realism.	No.	In the first study, we measured task performance, not use intentions.
Chapter 3	Aesthetics	Ethics, Aesthetics, Realism, Similarity	Yes. Perceived aesthetics had a larger impact on engagement than perceived realism.	No. Use intentions were slightly higher with an ugly than with a beautiful obstructing agent (trend).	
Chapter 4	(Facial) Similarity	Similarity	Yes. Facial similarity increased the involvement of all participants with an aiding agent, not an obstructing agent.	Yes (males), No (females). Males' use intentions were higher with a dissimilar than with a similar obstructing agent.	Facial similarity effects occurred unconsciously.
Chapter 5	(Body size) Similarity, Idealness (cf. Aesthetics)	Ethics, Idealness (cf. Aesthetics), Similarity	Yes. Perceived ethics had a larger impact on engagement than similarity and idealness.	Yes. Perceived ethics had a larger impact on use intentions than similarity and idealness.	The non-ideal (heavier) agents were perceived as ethically better than the ideal (slimmer) ones.
	Manipulations system-side	**User perceptions system-side**	**Effect on character-side?**	**Effect on system-side?**	**Remark**
Chapter 2	-	Relevance, Valence	Yes. Perceived relevance had a larger impact on engagement than realism.	Not tested.	
Chapter 3	Affordances	Affordances, Relevance, Valence	Yes. Affordances were more decisive for engagement than aesthetics. Relevance and valence were mediators in between affordances and engagement.	Yes. Affordances were more decisive for use intentions than aesthetics. Relevance and valence were mediators in between affordances and use intentions.	
Chapter 4	Affordances	Affordances, Relevance, Valence	Yes. Affordances were more decisive for engagement than similarity. Relevance and valence were mediators in between affordances and engagement.	Yes. Affordances were more decisive for use intentions than similarity. Relevance and valence were mediators in between affordances and use intentions.	
Chapter 5	-	Relevance, Valence	Yes. Relevance and valence were mediators in between ethics and engagement.	Yes. Relevance and valence were mediators in between ethics and use intentions.	

6.2 Main conclusions

The central question in this dissertation is how computer users respond to embodied agents. Based on the results presented above we provide the following answer. The system-side of user-agent interaction is more important for the various user responses (involvement, distance, and use intentions) than the appearance factors from the character-side (realism, aesthetics, facial similarity, and body shape similarity). The ethics factor is a noticeable exception, because, although it stems from the character-side, it is important for various user responses, including use intentions. People prefer not to interact with a bad or untrustworthy character. Remarkably, the affordances factor that stems from the system-side is more important for engagement than several typical factors from the character-side, such as aesthetics, realism and facial similarity, which all concern the agent's outer appearance. Appearance factors from the character-side may still affect user responses in unexpected ways, for example, in combination with the system-side factor affordances. When paired with obstructing affordances, positive features such as beauty and facial similarity may produce negative rather than a positive effects on user responses. Finally, users perceive an embodied agent within their task or goal context, in terms of (task-)relevance and valence. Relevance and valence serve as mediators in between effects of affordances and ethics, and user responses. User satisfaction, then, is a complex construct that depends on both the character-side and the system-side aspects.

We will elaborate further on five remarkable conclusions from the research: 1) realism is not that important for involvement, distance, and use intentions, 2) what is beautiful is not always used, 3) affordances and ethics are central factors for involvement, distance, and use intentions, 4) positive features do not always produce positive effects, and 5) relevance and valence of user goals are mediators. These conclusions and the implications of these conclusions for designers are discussed next.

6.2.1 Realism is not that important for involvement, distance, and use intentions

Many embodied agent studies and design efforts have focused on agent realism. They assume that user responses will be more positive, the more realistically embodied agents look and behave. Although designed and perceived realism did contribute to user involvement and distance in our studies, other user perceptions (e.g., of aesthetics, ethics, affordances, relevance) better explained user involvement and distance with embodied agents (Chapters 2 and 3). Thus, although realism contributed to involvement and distance, it was not the most important factor. In addition, perceived realism had no effect on use intentions. In other words, realistic agents were not likely to evoke higher or lower use intentions than unrealistic agents. These findings are in line with a recent meta-analysis by Yee et al. (2007), indicating that among many published papers, agent realism mainly affected subjective user responses, such as more positive social interactions, but not user behavior (e.g., use).

One could argue that the effect realism has on the user depends on the type of application in which an agent features. Certain applications may profit from highly realistic scenes and embodied agents. Certain applications aim to simulate or imitate the behavior of real people in various situations realistically, in order to analyze team behavior (Nair et al., 2004), to train people's negotiation skills (Core et al., 2006), and to train psychotherapists (Marsella et al., 2003). Even in such applications, however, not all aspects of realism may be equally important for engagement and use (cf. Hoorn et al., 2003). For example, realistic and believable behaviors (whether an agent acts like a real person) might be more important than form realism (whether an agent looks like a real person). In addition, even in simulations, factors such as aesthetics, ethics, and affordances are likely to affect the user. They do this in a complex way.

For designers, this means that attempts to design embodied agents to be as realistic as possible might not be as effective as hoped. Rather, other aspects of their designs, such as affordances and ethical considerations, should have priority over the design of realistic features. In addition, if graphical designers want to increase the level of user engagement, they should focus on aesthetic appeal rather than realism, because aesthetics is more important for engagement than realism.

6.2.2 What is beautiful is not always used

Well-known maxims in HCI are 'what is beautiful is good' (Dion et al., 1972), 'what is beautiful is usable' (Tractinsky et al., 2000), and 'attractive things work better' (Norman, 2002). These adages suggest that, in general, users expect a beautiful embodied agent to perform better and to be more trustworthy than an ugly one. As a result of such expectations, a beautiful appearance may positively affect first-time use, stimulate impulse purchases in a shop or affect the decision to download something from the Internet. For example, the high download rate of Virtual Katja[4], an embodied agent designed to search the Dutch Yellow pages, was not due to her search skills alone. The real Katja is a famous Dutch television actress and presenter, well-known for her good looks. Thus, designers may create beautiful-looking embodied agents or interfaces in order to persuade people to buy or try out a system.

Although agent beauty can attract initial interest effectively, the advantage may not last long. In the studies this dissertation presented, in which use intentions were measured *after actually using* an embodied agent, aesthetics actually did *not* determine the user's intentions to use the agent again. This calls into question the value of recent trends in HCI, in which the role of design aesthetics is preeminent (e.g., Tractinsky, 2004). Of course, aesthetics matters. Aesthetics matters for initial attraction and can improve a product's commercial value and desirability

[4]http://www.katja-schuurman.com, Retrieved January 1st 2008

(ibid). However, the initial positive effects of beauty fade after actual use of the application, at which point other features of the application become more important. People will leave an application untouched if it is not useful to them, even if its outer appearance is very nice. In addition, people are likely to avoid using an application if they perceive it as untrustworthy. This may explain the later rejection of Virtual Katja, which people probably downloaded because of her good looks but turned out to be spyware[5].

The above reasoning is in line with models that postulate that the importance of any given piece of information (such as aesthetics) as a determinant of a judgment becomes less important as a perceiver takes more pieces of information into account (cf., Anderson, 1981; Bull and Rumsey, 1988; Eagly et al., 1991). As the amount of other information about targets (here, the affordances of embodied agents) increases, the effect of aesthetics is likely to become smaller. Thus it seems as though recent trends in HCI, or graphical design in general, have overestimated the duration of the effect of aesthetics. Graphical design cannot make up for a useless or unreliable application. Thus, a prerequisite for prolonged use of an embodied agent is not beauty, but rather high-quality affordances and morally good behavior.

For designers this means that an embodied agent that 'promises' ease of use and trustworthiness should meet those expectations at the functional level. Masking an agent's low quality affordances or 'bad' behavior by making it look good is not a recommended design strategy. After all, obstructive affordances decrease the intention to use the agent in spite of its good looks. Therefore, if embodied agents do not already have high quality affordances and ethically good behavior, designers should not be tempted to put *too* much effort into designing them to look beautiful Such a strategy of graphical design could backfire.

6.2.3 Affordances and ethics are central factors for involvement, distance, and use intentions

In the studies this dissertation presents, users wanted to get good advice or wanted to perform a certain task efficiently. In our studies, the agent applications with high-quality aiding affordances and that were 'morally good' were the most engaging and used. Affordances and ethics were central to human-agent interactions. In a way, this is new to the field of HCI. Even though considerations of trustworthiness are important in HCI, especially in areas such as e-commerce, the field generally has not considered a more general ethics factor (is a system 'good' or 'bad').

Interestingly, however, the prevalence of ethics in the embodied agent domain is consistent with previous work on film and theater characters, which has demonstrated that ethics is more important than aesthetics and realism (e.g., Zillman, 1994; Konijn and Hoorn, 2005). As stated earlier in this dissertation, engagement

[5]http://www.nu.nl/news/514923/50/Commotie_om_'Virtuele_Katja'.html, Retrieved January 1st 2008

with film characters and engagement with embodied agents do have things in common. In addition, the system-side of I-PEFiC, the (perceived) affordances factor, also has great potential to affect user engagement (both involvement and distance). The affordances factor is even more essential to engagement with embodied agents than two of the 'engagement factors' from the character-side of I-PEFiC, aesthetics and realism. Furthermore, the affordances factor is also more essential to engagement than facial similarity, another character-side factor that deals with the agent's outer appearance. In general, most graphical design aspects are of *minor* importance for user engagement compared to functional and ethical design.

Similarly, affordances strongly affect intentions to use the agent, and of the typical 'engagement factors' from the character-side of I-PEFiC, only ethics has great potential to affect intentions to use the agent. In general, agents with obstructing affordances and agents that are morally bad will evoke tendencies toward non-use. Users do not want to interact with unhelpful and untrustworthy agents if they need to complete a task efficiently and effectively. Again, the graphical design is of lesser importance (Chapters 3 and 4).

One could argue that affordances and ethics comprise a single factor. Both affordances and ethics refer to qualities of an embodied agent that are either 'ok' (competent/ knowledgeable and trustworthy/ credible) or 'not ok' (incompetent/ dumb and untrustworthy/ incredible). However, despite this commonality, we think affordances and ethics should be regarded and treated as separate factors. Not only did the psychometric analyses performed in Chapter 3 (section 3.3.4) and Chapter 5 (section 5.2.1) indicate that affordances and ethics are separate factors, but there is also a conceptual difference between the two. Affordances have to do with the means a system provides to the user to perform a specific *task*, whereas ethics has to do with more *general* goals, such as maintaining privacy and feeling secure (Maslow, 1943). Whereas badly designed affordances may hinder the user's task, untrustworthy agents may decrease the user's sense of privacy and security. An agent can be designed with affordances that users perceive as helpful for their tasks but also be designed with spying behavior (the Virtual Katja is an example of such an agent) or deceptive clues (for example, an implementation of the GRETA agent by Rehm and André, 2005). In addition, an agent can have the best intentions, have no intentions to deceive the user, nor have a hidden agenda, and at the same time include affordances that the user deems useless and obstructing (Microsoft's Clippy is an example of such an agent).

Whether perceptions of affordances or perceptions of ethics are the most important may depend on the type of application. High quality affordances are especially important for agents that aim to help the user perform a task more efficiently (cf. Chapter 3), as well as for recommender systems, because users want high quality advice (cf. Chapter 4). Ethics is especially important when agents deal with personal interactions, such as e-health systems in which a user interacts with a care giver on personal issues (cf. Chapter 5). Then, special care is needed to design agents with features that are perceived as 'morally good'.

6.2.4 Positive features do not always produce positive effects

Usually, positive features produce positive effects. A beautiful agent will evoke more involvement than an ugly agent, and an agent with a face similar to the user will, most of the time, evoke more involvement than an agent with a dissimilar face (Chapters 3 and 4). Sometimes, however, positive features produce negative effects, and negative features produce positive effects. For example, a health advisor with a non-ideal, heavier body may be perceived as better morally (e.g., trustworthy) than ideally thin advisors, and consequently, the heavier agents evoke more involvement and higher use intentions. This effect might have been due to the specific (health) context (Chapter 5). Or, when an agent's affordances obstruct the user's task, a beautiful or facially similar agent may evoke less intentions to use it than an ugly or dissimilar agent (Chapters 3 and 4). Hence, the effect a feature has on the user depends on other features and the context.

For designers, this means that care in designing outer appearances of embodied agents is needed, because the most appropriate appearance may depend on other features or the context. Sometimes, it is better to design embodied agents that do not have a 'perfect' appearance. Test teams in a design group should investigate whether an agent's outer appearance affects user involvement, distance, and use intentions in the context for which the agent is designed, and in what way.

6.2.5 Relevance and valence of user goals are mediators

A large body of literature on user-centered design and related issues indicates the importance of considering user goals in system design. The user-centered approach assumes that, although design features shape user perceptions of and responses to a system, user goals do so even more. People use systems to achieve a goal, and obviously this is also true for embodied agent systems (e.g., Catrambone et al., 2004). Thus, the user-centered philosophy stresses that HCI practitioners should understand *to what end* a user wishes to use a product before implementing any code. Potential user goals may concern 1) how users want to *feel*, for example, have fun, feel connected with friends, feel like a hero, feel secure, 2) what users want to *do*, for example, obtain good advice, find music that they'll love, get the best deal, perform a productivity task (e.g., write a manuscript), and 3) what users want to *be*, for example, be cool, be popular, be respected by their peers, be a connoisseur of some sort . . . (e.g., Cooper et al., 2007). Designers will create more appropriate and satisfactory designs if they understand what computer activities mean to users in terms of their goals (cf. Cooper et al., 2007; Desmet, 2002).

The importance of goals for user responses is a key assumption, based on design cases, experience in the field, and common-sense reasoning. This dissertation adds to this knowledge in several ways. First, it provides a theoretical contribution. Based on theories from psychology, it suggests that not only task relevance but also task *valence* matters for user responses (cf. Hoorn et al., 2007). For example, an embodied agent may be relevant for the user's task goal but still not used. Users' outcome expectations might be negative, despite relevance. Think of Virtual Katja who provided users with the exact information they needed, but carried spyware at the same time. So yes, Virtual Katja was relevant to user goals - to find information - but raised negative outcome expectations as well - users expected trouble with respect to their privacy. An embodied agent may also be irrelevant for a user's main goal, for example efficient task completion, but he or she uses it anyway. For example, Clippy may often give useless tips but running the animation may still be fun. However, the experimental studies in this dissertation did not explore complex scenarios in which users had more than one goal, such as efficient task completion *and* enjoyment. Rather, we instructed participants to focus on one particular goal, for example, to perform a task quickly (Chapter 3) or to answer the questions the best they could (Chapter 4). Future research may inspect the roles of relevance and valence in complex scenarios, which are likely to exist in real-world settings.

The studies in this dissertation have shown that perceptions of relevance and perceptions of valence are important predictors of both user involvement, distance, and use intentions, which is in line with general expectations. More importantly, however, the studies have also shown that both relevance and valence merely acted as *mediators* in between (perceptions of) affordances and ethics on the one hand, and the various user responses on the other hand (see Table 6.1). This was unexpected because the user-centered design perspective emphasizes relevance and valence as a source of all user responses - not as factors that merely 'modify' effects. Mediators tell us more about the underlying mechanisms of effects. In this case, user's perceptions of affordances and ethics affect their perceptions of relevance and valence, which subsequently affect the various user responses (use intentions, involvement, distance).

Thus, after the user 'encodes' an agent in terms of affordances and ethics, it is checked for user goals. The user plays an active role during this process of goal comparison. Hence, this process is user-controlled. Users deliberately think about the relevance and valence of the agent for their goals. On the other hand, results showed that relevance and valence were did not mediate the relationship between aesthetics or facial similarity and the various user responses (Chapters 3 and 4). Thus users do not seem to evaluate appearance factors or graphical design factors in terms of their goals. They evoke direct and sometimes unconscious effects on user responses. The user has little or no control over this process, which can therefore be considered automatic. Thus, users may react to embodied agents consciously (controlled) or unconsciously (automatic) (cf. Schneider and Shiffrin, 1977).

Further, this dissertation makes a methodological contribution. The studies presented in this dissertation *empirically* tested the assumption that goal relevance and valence are important for user responses by means of structured questionnaires. An important assumption in the field of HCI is that user goals relate to engagement and use. However, empirical verification was still missing. Commonplace HCI methods (e.g., to determine requirements) aim to consider users and their goals. Typical approaches include task analyses, (open-ended) questionnaires, observations, focus groups, and interviews. The use of structured questionnaires allowed users to rate the agent in terms of relevance, valence, and other factors more systematically. This method allowed a more thorough inspection of the interrelations between the various factors that potentially contribute to user responses to embodied agents. Although construction of the questions may take some effort, structured questionnaires are convenient for user tests, both offline (e.g., using paper-and pencil) and online.

For HCI practitioners and designers, the above discussion means that, throughout an iterative design process, considerable effort should go into understanding both goal relevance *and* valence. First, designers should determine to what end users may wish to use an agent, and what negative outcome expectations may prevent the user from doing so. User tests may focus on whether they perceive the systems as relevant for (one or several of) their goals, and perhaps more importantly, what positive and negative outcomes users expect from using the system (e.g., get the advice they need, be entertained, (in)efficient task completion, feeling (in)secure). In addition, designers should understand what aspect(s) of an agent cause users to perceive the agent as irrelevant and/or negatively valenced. For example, is the code implemented so poorly that the agent offers no aiding affordances? Or, is the agent perceived as untrustworthy? Note that a beautiful, similar, or realistic appearance is unlikely to affect the user's outcome expectations. Designers can also ask users which outcome expectations are likely to affect their intentions to use the system. Do people use an agent because they expect it to provide *added* value compared to a situation in which no agent, or a different agent, is used to reach the same user goal? For example, do they expect to fulfill a task more successfully or efficiently, or do they use the agent because they expect to be entertained?

6.3 Theoretical implications: revised I-PEFiC

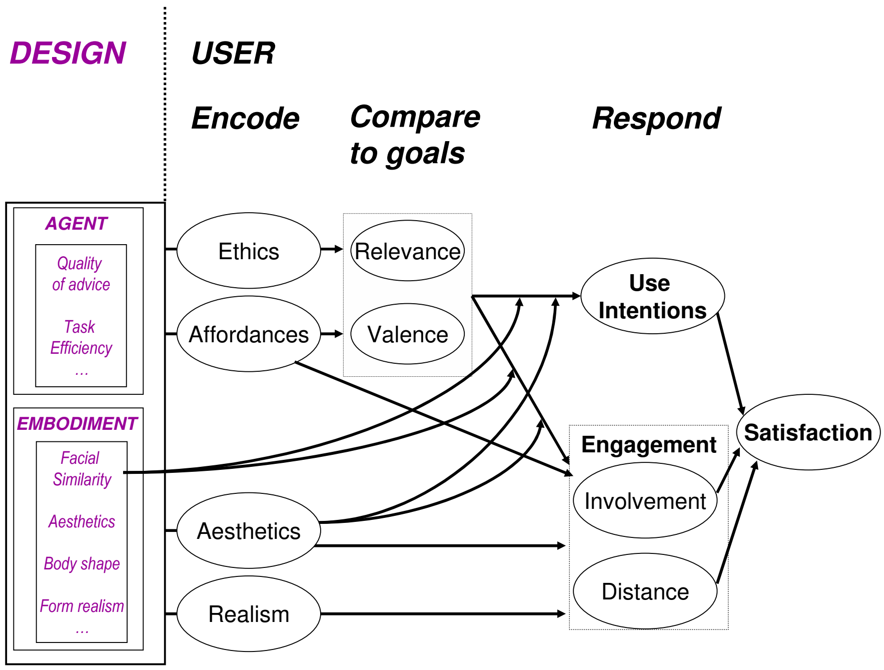
Now that a range of empirical studies has been performed, we can evaluate and adjust the initial theoretical model to describe the psychological processes that are involved in human-agent interactions more accurately. To disentangle user satisfaction, I-PEFiC consisted of two main building blocks: an interaction process, which perceptions of affordances evoked, and an engagement process, which perceptions of ethics, aesthetics, and realism.evoked The model includes relevance, valence and similarity as mediators in between both processes. The empirical data allowed us to refine these theoretical assumptions.

First, as expected, both engagement and use intentions contribute to user satisfaction. Second, both affordances and ethics play a major role in both engaging with and intending to use embodied agents. Relevance and valence are mediators in between these processes. Third, facial similarity, aesthetics and realism contribute to engagement, realism being the least important factor of the three and aesthetics coming in at second place. Unlike aesthetic and realism effects, facial similarity effects may occur without the participants consciously perceiving the similarity.

Fourth, aesthetics and facial similarity may also contribute to use intentions, but always in interaction with affordances. Realism, however, does *not* contribute to use intentions. Relevance and valence do not mediate between the various appearance factors and engagement and use intentions.

The revised I-PEFiC model is depicted in Figure 6.1, and will be explained next.

Figure 6.1: Revised I-PEFiC.



6.3.1 Design features

The left box of revised I-PEFiC represents design features of an embodied agent. The model distinguishes between two main categories of agent characteristics. The first category includes 'agent' features that deal with the agent's functionality, 'artificial intelligence', or 'mental' capacity. These features are usually designed by AI programmers. The efficiency with which an agent is able to perform a task depends on how the underlying software is designed, and the quality of advice that a virtual advisor gives depends on the way a reasoning engine of an agent is programmed. If the agent is a virtual teacher, then a cognitive model may give the agent the ability to teach and explain subject matters. If the agent is an empathic companion, then an emotional model may give the agent the ability to understand and respond to user emotions. Intelligent software functions such as cognitive models, emotion models, personality models, speech models, etc. are all part of the upper left, 'agent' category of revised I-PEFiC. Thus this category is closely related to the term 'agent' that is traditionally used in computer science, pointing to software with 'intelligent' capabilities (Wooldridge and Jennings, 1995).

The second category concerns 'embodiment' features that graphical designers usually determine. The embodiment of an agent can be designed in different ways, for example, with facial features similar to the user, to be beautiful or ugly, have a slim or heavy body shape, with realistic or unrealistic looks, and other qualities of appearance. Other features that deal with the agent's embodiment also pertain to this category, such as what parts of the body are present (head, head plus neck, torso, full body), the dimension (which may vary from simple 2D to complex 3D), hand and body gestures and movements, and facial and gaze expressions (cf. Ruttkay et al., 2004, p32-34).

The two categories generally are not completely separate from one another. That is, the role of an agent may be presented by both 'agent' and 'embodiment' features. For example, the embodied agent on the IKEA web site[6] does not only have certain (limited) capacities to help online IKEA customers, but is also dressed in yellow and blue like a real-life IKEA employee. Similarly, a virtual doctor that gives health advice may also look like a real doctor. Second, models designed at the AI level may specify how the agent should behave physically, on the graphical level. For example, the output of an emotion model is an emotional representation to be displayed by the virtual human. The thus selected 'emotion' (output from the emotion model, e.g., sadness) is converted into appropriate graphical components, such as facial expressions, head nods, gestures, eye gaze, and body movement (e.g., Gratch and Marsella, 2001). Thus, the left box is more complex than the model may suggest. However, the main concern of the revised I-PEFiC model is not the design complexities, but the user. Everything other than the left box of revised I-PEFiC is concerned with how users perceive and respond to embodied agents, which may be designed in very different ways.

6.3.2 Encode phase

When users interact with an embodied agent, they encode it in terms of 1) affordances, 2) ethics, 3) aesthetics, and 4) realism. The position of the factor in the model is somewhat related to its importance in the model. Our research found that aesthetics is more important than realism, and affordances and ethics are more important than aesthetics and realism. The relative importance of factors may change, of course, depending on the context.

Affordances exists relative to design features and to the user. Given that users are able to interact with an agent (e.g., users have a keyboard that correctly responds to their key presses), features in the 'agent' design category, such as the quality of advice and task efficiency, strongly influence perceptions of affordances. An agent that is efficient and/or gives high-quality advice offers helpful affordances

[6] http://www.ikea.com, Retrieved January 1st 2008

in the eyes of the user, and its affordances are perceived as ‘aids’. An agent that is inefficient and/or gives low-quality advice offers obstructing affordances in the eyes of the user, and its affordances are perceived as ‘obstacles’. Features of the ‘agent’ design category also influence perceptions of ethics (is the agent good or bad, trustworthy or untrustworthy). If an agent is designed to have only good intentions, the user is likely to perceive the agent as morally good. If the agent is designed as spyware, the user is likely to perceive the agent as untrustworthy or morally bad.

Features in the ‘embodiment’ design category affect perceptions of both aesthetics and realism, between others. People are likely to perceive a 3D virtual human with advanced body and face movements as more realistic than a 2D cartoon-like and static agent. Many (Western) users perceive agents with slim bodies, such as those of super models, as more beautiful than agents with more voluminous bodies. Thus, generally, designed features affect user perceptions in rather straightforward ways.

However, revised I-PEFiC recognizes that no one-to-one relation between design features and perceptions exists, as no lines are drawn between specific design features and specific user perceptions in Figure 6.1. First, features from not only the ‘agent’ design category but also the ‘embodiment’ design category may affect user perceptions of affordances and ethics. For example, the body shape of an agent may affect perceptions of ethics (cf. Chapter 5). In addition, features from not only the ‘embodiment’ design category, but also the ‘agent’ design category may affect user perceptions of realism. For example, the user may perceive an agent that gives very bad advice (an ‘agent’ feature) as unrealistic. Second, user perceptions do not only depend on the design, but on the user and the cultural context as well. For example, users may perceive an agent that uses speech to give high-quality advice to its users as having ‘aiding’ affordances. However, for users that are deaf, the affordance of communicating with the agent does not even exist. Or, whereas people may perceive a slim agent to be more beautiful than one that is bigger, the reverse may hold in other cultural contexts. In Morocco and other African countries, large girth is a sign of well-being and something to strive for (Rguibia and Belahsen, 2006). Thus, standards of beauty change through times and places, and this is true for moral standards (what is good and what is bad) as well. Third, sometimes a design feature does not affect user perceptions at all. For example, people may not consciously perceive designed facial similarity between an agent and themselves (cf. Chapter 4).

Although designs are usually tested with objective measures such as error rates, it is more important to understand how designs are perceived by the user. This is especially important because the intentions designers have often do not match user perceptions and responses (cf. Norman, 1988).

6.3.3 Compare phase

After encoding an embodied agent on various dimensions, users enter the phase in which they ‘compare’ the agent to their goals. That is, they perceive whether or not the agent is relevant to their goals (relevance), and whether they expect positive or negative outcomes of using the agent (valence). This is consistent with psychological theories suggesting that human action is goal-driven, to a considerable extent (e.g., Gollwitzer and Bargh, 1996).

When the user’s goal is to complete a task efficiently and effectively (as was the case in the studies presented in this dissertation), perceptions of relevance and valence typically depend on perceptions of affordances and ethics, and not on perceptions of aesthetics and realism. In general, an agent is perceived as relevant and positively valenced when users think it provides aiding (not obstructing) affordances, and/or when users think the agent is morally good (e.g., good intentions, trustworthy). In addition, positive perceptions of affordances and ethics generally lead to positive perceptions of relevance and valence. The arrows between perceived affordances/ethics and perceived relevance/valence represent positive relations.

However, it might be that, in contexts other than the ones presented in this dissertation, users perceive embodied agents as having aiding affordances (e.g., providing high quality advice), but also as irrelevant for whatever task the user is performing. In addition, the studies in this dissertation did not include situations in which users perceived an embodied agent as having aiding affordances but being untrustworthy at the same time. The (relative) importance users assign to each of their goals within a context may determine whether affordances or ethics influences user perceptions of relevance and valence more. For example, if users want to perform a financial transaction online, they might think it is more important that they can trust the system than that the transaction is performed quickly. Such a scenario may be used in future studies to study the relation between perceptions of affordances and ethics, and perceptions of relevance and valence more deeply.

In the initial, theoretical I-PEFiC model, perceptions of similarity were also expected to be central to the comparison phase. However, the studies in this dissertation did not provide evidence that perceptions of similarity were (consciously) affected by factors in the encode phase nor that perceptions of similarity affected the subsequent user responses. Therefore, revised I-PEFiC excludes perceived similarity as a factor in the compare phase. This does not mean that similarity between agent and user does not affect the user. Similarity between agent and user may affect user responses at an unconscious level, that is, not via user perceptions of similarity, but automatically, skipping the encode and compare phases.

6.3.4 Respond phase

Users may respond to embodied agents on behavioral, emotional, social, and cognitive (e.g., learning) levels. Of all possible user responses, the revised I-PEFiC model aims to explain involvement and distance (emotional approach and avoidance), use intentions (actual approach and avoidance), and final user satisfaction. This is the 'respond' phase in revised I-PEFiC.

Both 'agent' and 'embodiment' design affect use intentions, via user perceptions. First, perceptions of affordances and ethics (which are most strongly related to 'agent' design features) affect the individual's intentions to use an embodied agent again in the future. Perceptions of relevance and valence play mediating roles herein. Second, facial similarity may affect use intentions in interaction with affordances. When an agent has helpful affordances, facial similarity may increase intentions to use the agent again. When an agent has obstructing affordances, facial similarity may decrease intentions to use the agent again. In certain contexts, the interaction effect may only occur for males, not females.

Third, the design factor aesthetics, which is closely related to user perceptions of aesthetics, may also affect use intentions in interaction with affordances. When an agent has helpful affordances, aesthetics may increase intentions to use the agent again. When an agent has obstructing affordances, aesthetics may decrease intentions to use the agent again.

Both 'agent' and 'embodiment' design also affect involvement and distance. First, (perceived) affordances and ethics affect involvement and distance. Relevance and valence play mediating roles between ethics and involvement, ethics and distance, and affordances and distance. Affordances affect involvement directly; that is, relevance and valence do not play mediating roles herein. Second, facial similarity affects involvement in interaction with affordances. It boosts involvement only when the agent has helpful affordances, not when an agent has obstructing affordances. Third, aesthetics and realism affect involvement and distance. The more beautiful and realistic, the more involved and the less distant a user feels with the agent.

User satisfaction can be seen as an end-state of a user after the interaction with an agent has finished. Satisfaction depends on both emotional and actual approach and avoidance, and can thus be considered a complex construct. A user's satisfaction with an embodied agent does not say much about processes at the psychological level.

In the response phase, the two dimensions of engagement are involvement *and* distance, next to each other. The research in this dissertation has demonstrated that involvement and distance are two separate dimensions and not the ends of one continuum (cf. Konijn and Hoorn, 2005; Hoorn and Van Vugt, 2006; Konijn et al., 2007; Konijn and Bushman, 2007). However, it is difficult to specify more accurately under what conditions a feature affects involvement, or distance, or both, and to what extent. Whereas beauty increased involvement in an efficiency context, ideal similarity (which is related to beauty) increased distance in an e-health context. A clear pattern remained absent throughout the studies. Therefore, we *do* know it is insufficient to measure either involvement or distance but we do not know what affects each factor. Thus, the revised I-PEFiC model merely specifies that a certain feature may affect engagement, but not which dimension it affects.

Future endeavors may try to disentangle the two dimensions of engagement. In addition, future endeavors may focus on the relation between engagement and use intentions, that is, between emotional approach (involvement) and avoidance (distance) and actual approach (use) and avoidance (non-use). For example, under what conditions does emotional approach lead to actual approach, and under what conditions does it not?

6.4 Future directions

People may interact with specific embodied agents only once. For example, a virtual museum guide is typically used by museum visitors who visit a museum once. With embodied agents that appear on the Internet and in office applications, people may interact more often. At least, that is what designers hope. It is not yet clear how user experiences develop over time. It is likely that experiences of using an agent feed back on the user's mental model of that agent, creating a reactive system. For example, people might have learned what affordances belong to a particular agent, shaping their perceptions accordingly. (Updated) mental models may co-determine the relevance and valence people attribute to the agent in view of their goals. Long-term investigations might tell how feedback loops actually work and how user perceptions develop over time. Further, aesthetic aspects may matter more for initial attraction, and hence, for novice users, than after use. Whether the effects of appearance factors such as aesthetics eventually disappear, or whether they remain present, is yet unclear. Do experienced users only care for the agent's affordances and trustworthy behavior, or are they still affected by the agent's beautiful design? How long do first impressions linger on? Future research may thus focus on the effects of first impressions on subsequent interactive experiences and behavior. In addition, longitudinal research may study the importance of the various factors over time.

In such longitudinal studies, it is important to realize that users may develop habits over time. Habits may drive people to use an agent even if that behavior serves no particular goal. Users may also disregard agents due to sheer habit. In the case of habits, users repeat acts without thinking consciously about what they do (e.g., Aarts and Dijksterhuis, 2000). Thus, although much of human action is goal-driven, sometimes it is not. In the studies we conducted, use intentions largely were based on conscious processes (such as the comparison of agents to user goals). Nevertheless, unconscious processes may also trigger use intentions and actual use (as was the case with the facial similarity effects in Chapter 4). The role of conscious versus unconscious processes and habit in embodied agent use (and technology use in general) is an unexplored but highly interesting area of research that may be important for understanding the success or failure of embodied agents in the real world.

To increase ecological validity, future work may consider more complex, natural scenarios. In the real world, users have a range of different characteristics, personalities, beliefs and attitudes, they may have multiple goals and work in a variety of (task-)contexts, and they are affected by prevailing social norms and other societal factors. A challenge for future research is to study the effects that such factors may have on user perceptions of and responses to embodied agents. For example, users seeking fun may think the visual appearance of embodied agents is important for enjoyment. Therefore, appearance features may affect their use intentions in entertainment applications to a larger extent than in, for example, productivity applications. In addition, unlike productivity applications, obstacles and morally bad behavior may positively contribute to enjoyment in games (e.g., Song and Lee, 2007; Pagulayan et al., 2003). Perhaps it is for this reason that the highly popular game World of Warcraft features two main agent types, members of the alliance or horde, which represent the good (alliance) and the bad (horde). It is not yet known to what extent visual aspects, affordances, and ethical considerations will affect users of entertainment and other types of applications. In addition, some of the above factors may be more important for engaging with an embodied agent, and others for use intentions and actual use. For example, user behavior may depend to a larger extent on societal factors and perceptions of self-efficacy than emotional reactions. Revised I-PEFiC may be used as the theoretical framework for studies on such topics. Applying revised I-PEFiC to a variety of applications and contexts may lead to a more complete and accurate picture of the psychological processes human-agent interactions evoke.

The revised I-PEFiC model may also be used to study human-robot interactions (cf. Bosse et al., 2008). Both embodied agents and robots have 'humanoid' features, for example with regard to their appearance. In addition, embodied agents and robots are both 'tools' that can be used to accomplish tasks, and with which people can interact. Hence, as with embodied agents, appearance factors such as aesthetics, facial similarity and realism, as well as the factors ethics and affordances, are likely to play a role in engaging with and/or using a robot. Future work may examine the commonalities and differences between human-agent and user-robot interactions.

Future studies may further uncover how self-reported engagement and use intentions relate to actual behavior and use of embodied agents. To study actual use, one could create an environment in which users have the possibility to turn the embodied agent on and off. In addition, future studies may uncover different types of effects that are not covered by revised I-PEFiC and how these relate to embodied agent engagement, use, and satisfaction. One may think of cognitive and social effects such as mimicry, remembrance, learning, task performance, job satisfaction, advice following, and medication intake. It might be the case that certain cognitive and social effects are most likely to manifest themselves when users are involved with an embodied agent, as we speculated in Chapter 2.

Applying person perception and HCI principles to the embodied agent domain helps us to better understand user responses to embodied agents. To understand the nature of human-agent interactions at a more profound level, it would be interesting to investigate human-agent interactions from a neurological perspective. Only recently have scientists started to study the phenomenon of 'anthropomorphism' from a neuro-scientific perspective. That is, Harris and Fiske (2008) wondered whether people respond to objects in a similar way as to people by investigating brain activation areas. Interestingly, they showed that *different* parts of the brain are activated depending on whether humans perceive people or objects such as pens and paper. Yet, although some studies have started to investigate the interaction of humans and virtual figures from a neurological perspective (Schilbach et al., 2006), unclear is what precisely happens on the neurological level when people perceive embodied agents - objects that look like humans. Based on the results of the current dissertation, one may expect that in human-agent interactions, both the parts related to person perception *and* the parts related to object perception are activated at the same time.

This dissertation started with the idea that human responses to embodied agents depend on both their function and appearance, as is the case with elephants. This appeared to be a valid assumption but we can now add to this knowledge. An important difference between elephants and agents is that, whereas in the animal world, the prettiest is most likely to survive in the end (cf. 'Survival of the prettiest' Etcoff, 1999), the prettiest may not survive in the agent world. The more beautiful animal is also usually the animal that is stronger, and therefore better equipped for survival. In contrast, the more beautiful agent is not always the better agent. An agent's outer appearance is unrelated to the quality of the actions it can perform. It is probable that, because the outer appearance of an agent does not tell much about how well it can perform a task, users just do not care as much about it as they would in the real world. This is a noteworthy difference between the real world that elephants live in versus the virtual world of agents.

Bibliography

Aarts, H. and Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology*, 78:53–63.

Anderson, N. H. (1981). Integration theory applied to cognitive responses and attitudes. In Petty, R. E., Ostrom, T. M., and Brock, T. C., editors, *Cognitive responses in persuasion*, pages 361–397, Hillsdale, NJ.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing source. *Psychological Review*, 88:354.

Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. Yale University Press, New Haven, CT.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., and Bailenson, J. (2002). Immersive virtual environment technology as a research tool for social psychology. *Psychological Inquiry*, 13:103–125.

Bosse, T., Hoorn, J. F., Pontier, M. A., and Siddiqui, G. F. (2008). A robot's experience of another robot: Simulation. Submitted.

Brosnan, M. (1998). *Technophobia*. Routledge, London.

Bull, R. and Rumsey, N. (1988). *The social psychology of facial appearance*. Springer-Verlag, New York.

Catrambone, R., Stasko, J., and Xiao, J. (2004). ECA as user interface paradigm. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 239–270, Dordrecht / Boston / London. Kluwer.

Cerrato, L. and Ekeklint, S. (2004). Evaluating users reactions to human-like interfaces: Prosodic and paralinguistic features as new evaluation measures for users satisfaction. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 101–124. Kluwer Academic Publishers.

Cooper, A., Reimann, R., and Cronin, D. (2007). *About Face 3: The Essentials of Interaction Design*. John Wiley & Sons, Indianapolis, Indiana, USA.

Core, M. G., Traum, D., Lane, H. C., Swartout, W., Marsella, S., Gratch, J., and van Lent, M. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82:685–701.

Dehn, D. M. and Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1):1–22.

Desmet, P. (2002). *Designing emotions*. PhD thesis, Delft University of Technology, Delft.

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3):285–290.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., and Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110:109–128.

etcoff, N. (1999). *Survival of the prettiest. The science of beauty*. Anchor Books, New York.

Gollwitzer, P. M. and Bargh, J. A. (1996). *The psychology of action: Linking cognition and motivation to behavior*. Guilford, New York.

Gratch, J. and Marsella, S. (2001). Tears and fears: modeling emotions and emotional behaviors in synthetic agents. In Müller, J. P., Andre, E., Sen, S., and Frasson, C., editors, *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 278–285, Montreal, Canada. ACM Press.

Gratch, J. and Marsella, S. (2005). Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, 19(3-4):215–233.

Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors (2006). *Proceedings of Intelligent Virtual Agents 2006, IVA'06, Marina del Rey, CA, USA, August 21-23, 2006*, volume 4133 of *Lecture Notes in Artificial Intelligence*. Springer.

Han, S. and Hong, S. (2003). A systematic approach for coupling user satisfaction with product design. *Ergonomics*, 46(13/14):1441–1461.

Harris, L. T. and Fiske, S. T. (2008). Neural correlates of anthropomorphizing objects. *Social Cognition*. In press.

Hoorn, J. F. and Konijn, E. A. (2003). Perceiving and experiencing fictional characters: An integrative account. *Japanese Psychological Research*, 45(4):221–239.

Hoorn, J. F., Konijn, E. A., and Van der Veer, G. C. (2003). Virtual reality: Do not augment realism, augment relevance. *Upgrade - Human-Computer Interaction: Overcoming Barriers*, 4(1):18–26. Retrieved February 15, 2003, from http://www.upgrade-cepis.org/issues/2003/1/up4-1Hoorn.pdf.

Hoorn, J. F., Konijn, E. A., Van Vliet, H. C., and Van der Veer, G. C. (2007). Requirements change: Fears dictate the must haves; desires the won't haves. *Journal of Systems and Software*, 3(80):328–355.

Hoorn, J. F. and Van Vugt, H. C. (2006). The role of social norm in user-engagement and appreciation of the web interface agent Bonzi Buddy. In Gratch et al. (2006), page 456.

Konijn, E. A. and Bushman, B. J. (2007). World leaders as movie characters? Perceptions of J.W. Bush, T. Blair, O. Bin Laden, and S. Hussein. *Media Psychology*, 9(1):157–177.

Konijn, E. A. and Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology*, 7(2):107–144.

Konijn, E. A., Nije Bijvank, M., and Bushman, B. (2007). I wish I were a warrior: the role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, 43(4):1038–1044.

Lindgaard, G. and Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15(3):429–452.

Marsella, S. C., Johnson, W. L., and LaBore, C. M. (2003). Interactive pedagogical drama for health interventions. In *Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies*, pages 341–348, Amsterdam. IOS Press.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50:370–396.

Nair, R., Tambe, M., Marsella, S., and Raines, T. (2004). Automated assistants to analyze team behavior. *Journal of Autonomous Agents and Multi-Agent Systems*, 8(1):69–111.

Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.

Norman, D. (2002). Emotion & design: attractive things work better. *interactions*, 9(4):36–42.

Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., and Fuller, T. (2003). User-centered design in games. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pages 883–906. Lawrence Erlbaum Associates.

Pelachaud, C., Martin, J., André, E., Chollet, G., Karpouzis, K., and Pelé, D., editors (2007). *Proceedings of Intelligent Virtual Agents 2007, IVA'07, Paris, France, September 17-19*, volume 4722 of *Lecture Notes in Artificial Intelligence*. Springer.

Rehm, M. and André, E. (2005). Catch me if you can – exploring lying agents in social settings. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 937–944, New York, NY, USA. ACM Press.

Rguibia, M. and Belahsen, R. (2006). Body size preferences and sociocultural influences on attitudes towards obesity among Moroccan Sahraoui women. *Body Image: An International Journal of Research*, 3(4):395–400.

Ruttkay, Z., Dormann, C., and Noot, H. (2004). Embodied conversational agents on a common ground. A framework for design and evaluation. In Ruttkay, Z. and Pelachaud, C., editors, *From brows to trust: Evaluating embodied conversational agents*, pages 27–66. Kluwer Academic Publishers.

Saffer, D. (2007). *Designing for interaction. Creating smart applications and clever devices*. New Riders.

Sanders, G. A. and Scholtz, J. (2000). Measurement and evaluation of embodied conversational agents. In Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors, *Embodied Conversational Agents*, pages 346–373. MIT Press.

Schilbach, L., Wohlschlaeger, A., Krämer, N. C., Newen, A., Zilles, K., Shah, J. N., Fink, G. R. and Vogeley, K. (2006). Being With Virtual Others: Neural Correlates of Social Interaction. *Neuropsychologia*, 44:718–730.

Schneider, W. and Shiffrin, R. M. (1977). Controlled and automatic human information processing i: Detection, search, and attention. *Psychological Review*, 84:1–66.

Song, S. and Lee, J. (2007). Key factors of heuristic evaluation for game design: Towards massively multi-player online role-playing game. *International Journal of Human-Computer Studies*, 65(8):709–723.

Tractinsky, N. (2004). Towards the study of aesthetics in information technology. In *Proceedings of the 25th Annual International Conference on Information Systems*, pages 771–780.

Tractinsky, N., Shoval-Katz, A., and Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13(2):127–145.

Venkatesh, V. (2000). Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4):342–365.

Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A. (1997). Paradise: A framework for evaluating spoken dialogue agents. In *Proceedings of the 35th ACL and 8th EACL*, pages 271–280, San Francisco, CA. Association for Computational Linguistics and Morgan Kaufmann.

Wooldridge, M. J. and Jennings, N. R. (1995). Intelligent agents: theory and practice. *Knowledge Engineering Review*, 10(2):115–152.

Yee, N., Bailenson, J. N., and Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, pages 1–10, New York, NY, USA. ACM Press.

Zillman, D. (1994). Mechanisms of emotional involvement with drama. *Poetics*, 23:33–51.

Glossary

(Perceived) Aesthetics The degree to which users think the agent has a beautiful or ugly appearance.

(Perceived) Affordances The degree to which users perceive the agent's potentials for action as aids or obstacles. Agents are likely to be perceived as 'aids' when they are efficient and/or give high-quality advice. Agents are likely to be perceived as 'obstacles' when they are inefficient and/or give low-quality advice.

Distance A felt tendency to psychologically avoid the agent. The concept of distance covers a broad range of neutral to negative affects toward an agent, such as antipathy, aversion, irritation, boredom.

Engagement Engagement with an agent consists of parallel tendencies to approach (involvement) and avoid (distance) the agent. The two dimensions of engagement (involvement and distance) are unipolar processes and run in parallel.

(Perceived) Ethics The degree to which users think the agent is morally good or bad. Depending on the context, a 'good' agent may be trustworthy, credible, friendly, and have good intentions. A 'bad' agent may be untrustworthy, incredible, unfriendly, and have bad intentions such as spying behavior or deception.

Involvement A felt tendency to psychologically approach the agent. The concept of involvement indicates the level of psychological investment in an agent, and covers a broad range of neutral to positive affects toward an agent, such as sympathy, empathy, challenge, identification, immersion.

(Perceived) Realism The degree to which users think the agent resembles a real person or animal in form and behavior.

(Perceived) Relevance The degree to which users think the agent is important for their goals. Thus, relevance reflects the personal importance attached to the agent. Relevance intensifies the affective reaction toward the agent.

Satisfaction A subjective sum of several user experiences with the agent, such as involvement, distance and use intentions.

(Perceived) Similarity The degree to which users recognize themselves in an agent. Recognition of the self in an agent can occur on dimensions such as face, gender, ethnicity, personality, and attitude.

Use intention A felt tendency to use, or not to use, the agent. Use intention is a bipolar construct.

Use Actual usage, such as asking the agent for advice.

(Perceived) Valence The degree to which users expect that interacting with the agent results in a positive or negative outcome. Valence reflects the direction (positive, neutral, negative) of the affective responses based on the implied outcome of the encountered event.

Samenvatting (Dutch summary)

'Embodied agents' vanuit een gebruikersperspectief

In dit proefschrift staat de vraag centraal hoe computergebruikers reageren op zogenaamde *embodied agents*. Embodied agents zijn computerprogramma's die op het computerscherm verschijnen met een mens- of dierachtige verschijning. De virtuele IKEA-medewerker Anna[7] of Virtuele Katja[8] op het Internet, de paperclip in programma's van Microsoft en karakters in virtuele werelden zoals Grand Theft Auto zijn voorbeelden van embodied agents. Echter, het is onduidelijk waarom computergebruikers zich emotioneel binden aan, betrokken voelen bij of juist afstand voelen tot een embodied agent en waarom zij een agent wel of niet willen gebruiken. Om meer te weten over de wijze waarop betrokkenheidsprocessen en interactieprocessen zich voltrekken, is het van belang om embodied agents vanuit een psychologische invalshoek te onderzoeken. Oftewel, embodied agents dienen vanuit het perspectief van de gebruiker onderzocht te worden als we meer te weten willen komen over hoe computergebruikers reageren op zulke embodied agents.

Het doel van dit proefschrift is dan ook om meer inzicht te verschaffen in de psychologische processen die worden opgewekt als een embodied agent op het scherm verschijnt ten einde die ontwerpen te verbeteren. Om deze processen nader te onderzoeken, is gebruik gemaakt van verschillende theorieën. Ten eerste werd gebruik gemaakt van het PEFiC model (Perceiving and Experiencing Fictional Characters), dat het betrokkenheidsproces verklaart van film- en televisiekijkers op fictieve personages, zoals Dracula en Superman. PEFiC stelt dat een complex geheel van factoren zoals ethica (goed versus kwaad, betrouwbaar versus onbetrouwbaar), esthetica (mooi versus lelijk), realisme, gelijkenis (bijvoorbeeld in uiterlijk of gedrag), persoonlijke relevantie, en valentie (positieve of negatieve uitkomstverwachtingen) bijdragen aan de emotionele binding met fictieve personages, oftewel aan het betrokkenheidsproces. In dit proefschrift is het PEFiC model

[7] http://www.ikea.com
[8] http://www.katja-schuurman.com

aangepast aan het embodied agent domein. Embodied agents, in tegenstelling tot fictieve karakters, hebben een interactieve component. Gebruikers kunnen actief met embodied agents interacteren en gebruikmaken van embodied agents bij de taakuitvoering, zoals bij het zoeken naar informatie op het Internet. Het PEFiC model is daarom uitgebreid met een interactieproces en wordt derhalve Interactive PEFiC, oftewel I-PEFiC, genoemd. Dit interactieproces is gebaseerd op verschillende theorieën uit de literatuur die verklaren waarom computergebruikers systemen van allerlei aard wel of niet gebruiken. Het interactieproces stelt dat de gebruiksmogelijkheden die een embodied agent aan de gebruiker verschaft (affordances) bepalen of een gebruiker een embodied agent wel of niet wil gebruiken. Het stelt verder dat de relevantie van de agent voor de taak van de gebruiker en de uitkomstverwachtingen die de gebruiker heeft (bijvoorbeeld, is de agent effectief en efficient?) hierbij ook een rol spelen. I-PEFiC veronderstelt verder dat gebruikerstevredenheid afhangt van zowel het betrokkenheidsproces als het interactieproces. I-PEFiC bestaat dus uit twee lagen. De eerste laag heeft te maken met de karakterkant van embodied agents, en beschrijft een betrokkenheidsproces. De tweede laag heeft te maken met de systeemkant van embodied agents en beschrijft een interactieproces.

Ondanks dat het I-PEFiC model is gebaseerd op bestaande theorieën en modellen, dient het model nader, empirisch, onderzocht te worden. Er dient bijvoorbeeld onderzocht te worden of factoren die reacties van kijkers op fictieve personages in films kunnen verklaren, ook werkelijk reacties van computergebruikers op embodied agents kunnen verklaren. Interessant is verder of de typische karakterfactoren (zoals esthetica en realisme) ook het interactieproces of de systeemkant van mens-agent interacties beïnvloeden en of de typische interactiefactoren (zoals affordances, taakrelevantie) ook het betrokkenheidsproces of de karakterkant beïnvloeden. Om dit te onderzoeken, zijn in dit proefschrift verschillende empirische studies uitgevoerd. Er werden steeds één of meer factoren van een embodied agent gevariëerd om het (gecombineerde) effect op de emotionele binding, gebruiksintenties en tevredenheid te onderzoeken.

Tabel 1.1 in Hoofdstuk 1 geeft een overzicht van de gemanipuleerde en gemeten factoren per studie. In Hoofdstuk 2 vroegen wij ons af of (vorm) realisme een sleutelfactor is voor gebruikersreacties op embodied agents, of dat andere factoren belangrijker zijn. Resultaten lieten zien dat esthetica and taak-relevantie belangrijker waren dan realisme. In Hoofdstuk 3 vroegen wij ons daarom af hoe esthetica, een factor van de karakterkant van I-PEFiC, en affordances, een factor van de system- of taakkant van het model, gebruikersreacties op embodied agents beïnvloeden. De resultaten lieten een sterk effect zien van affordances binnen de systemkant, maar ook een sterk kruis-effect naar de karakterkant (zie ook Figuur 1.3). In Hoofdstuk 4 besloten we daarom de effecten van affordances op gebruikersreacties nader te bestuderen. We vroegen ons af hoe affordances en gezichtsgelijkheid tussen gebruiker en embodied agent, een andere factor uit de karakterkant van het model, gebruikersreacties beïnvloeden. Affordances hadden weer een groot effect binnen de systeemkant als ook een sterk sterk kruis-effect naar de karakterkant. Gezichtsgelijkheid had een effect binnen de karakterkant, en interessant genoeg, ook naar de systeemkant, maar alleen voor mannen. De effecten van gezichtsgelijkheid waren opmerkelijk, aangezien proefpersonen zich *niet bewust* waren van de manipulatie van gezichtsgelijkheid. In Hoofdstuk 5 vervolgden wij de studie naar gelijkheidseffecten, maar op een andere dimensie. We vroegen ons af of gelijkheid van lichaamsomvang en ideaalheid van lichaamsomvang (wat gerelateerd is aan de factor esthetica) invloed hebben op gebruikersreacties op embodied agents. Het bleek dat een niet-ideaal figuur een sterker effect had dan gelijkheid op zowel de karakterkant als de systeemkant, en dat deze effecten vooral indirect plaatsvonden, via percepties van ethica. Interessant was dat de niet-ideale (dikkere) agenten als ethisch beter (betrouwbaarder) werden gezien dan de ideale (dunnere) agenten.

Gebaseerd op deze resultaten, die ook gepresenteerd zijn in Tabel 6.1, kunnen we nu het volgende antwoord geven op de vraag hoe computergebruikers reageren op embodied agents. De systeemkant van mens-agent interactie is belangrijker voor de verschillende gebruikersreacties dan de factoren van de karakterkant die te maken hebben met het uiterlijk van de agent (realisme, esthetics, gezichtsgelijkheid, gelijkheid en ideaalheid van lichaamsomvang). De factor ethica is een opmerkelijke uitzondering. Ondanks dat de factor ethica behoort tot de karakterkant, is het belangrijk voor alle verschillende gebruikersreacties, dus ook voor gebruiksintenties. Een slecht karakter wordt liever niet gebruikt. Opmerkelijk is dat de typische interactiefactor affordances meer van belang lijkt te zijn voor emotionele binding dan een aantal typische karakterfactoren zoals esthetica, realisme en gezichtsgelijkenis, die alle met het uiterlijk van de agent van doen hebben. Uiterlijke factoren van de karakterkant kunnen gebruikersreacties wel op onverwachte manieren beïnvloeden, in combinatie met een factor van de systeemkant, affordances. Als affordances de gebruiker belemmeren om zijn doel te behalen, kan een positief kenmerk zoals schoonheid en gezichtsgelijkheid een negatief effect hebben op gebruikersreacties. Verder nemen gebruikers een embodied agent binnen hun taak- of doelcontext waar, in termen van taakrelevantie en valentie (verwachtingen of het gebruik van de agent wel of niet zal leiden tot taakvolbrenging). Een agent die wordt waargenomen in termen van affordances en ethica wordt 'getoetst' op taakrelevantie en valentie, en deze percepties van relevantie en valentie brengen vervolgens de verschillende gebruikersreacties teweeg (relevantie en valentie zijn zogenaamde 'mediatoren'). Gebruikerstevredenheid, ten slotte, hangt af van zowel het betrokkenheidsproces als het interactieproces, en is dus een complex construct. Dat een gebruiker tevreden is met een agent wil nog niet zo veel zeggen over wat er werkelijk in zijn of haar hoofd omgaat. Gebaseerd op het bovenstaande kunnen we een vijftal opmerkelijke gevolgtrekkingen onderscheiden: 1) realisme is niet zo belangrijk voor emotionele binding en gebruiksintenties, 2) wat mooi is hoeft nog niet gebruikt te worden, 3) affordances en ethica zijn centrale factoren voor emotionele binding en gebruiksintenties, 4) positieve kenmerken hoeven niet altijd tot positieve effecten te leiden, en 5) relevantie en valentie van gebruikersdoelen zijn mediatoren.

De resultaten van de studies hebben geleid tot de ontwikkeling van het Herziene I-PEFiC model (zie Figuur 6.1), dat op empirische basis de psychologische processen beschrijft die worden opgewekt als een embodied agent op het scherm verschijnt. Dit proefschrift heeft dus een bijdrage geleverd aan theorievorming rondom embodied agents. Natuurlijk kunnen deze inzichten ook gebruikt worden in de praktijk, door ontwerpers van embodied agents. Het is voor ontwerpers bijvoorbeeld van belang te beseffen dat de interactiecomponent van agents en de taken van gebruikers niet alleen van cruciaal belang zijn voor gebruiksintenties, maar dat ze zelfs ook de emotionele binding met agents sterk beïnvloeden, sterker dan het uiterlijk van agents. Een realistische of mooie agent is dus zeker niet zaligmakend. Een grafisch knap ontworpen uiterlijk kan ook niet compenseren voor een onhandige interface of slecht functionerende software. Het kan dan zelfs nog negatiever uitpakken dan bij een lelijke interface. Een effectieve designer maakt een agent eerst intelligent genoeg zodat de gebruiker er werkelijk iets aan heeft, om daarna pas als toegevoegde waarde het uiterlijk van de agent tot grotere schoonheid te brengen.

Acknowledgements

Thanks, first, to my supervisors and copromotors, Elly Konijn and Johan Hoorn. You have supported me in all my PhD activities, from coming up with new experimental designs, to carrying out the experiments, to analyzing the data, to writing papers and chapters, to attending conferences. I do not believe I could have received better guidance to complete the PhD work. I would also like to thank my promotors, Jan Kleinnijenhuis and Gerrit van der Veer, as well as the members of the reading committee, for their support and valuable feedback. Further, I am grateful to Jeremy Bailenson - you gave me the opportunity to work in the VHIL lab at Stanford University for several months.

I thank my colleagues and fellow PhD students at both the Communication Science and the Computer Science departments. With you, I discussed my work, I went to the PhD club, I went to conferences, I wrote papers, I shared a room, I had lunch, and some of you even participated in my experiments. I really appreciate all of that. I would also like to thank the students that helped me setting up and performing the experiments. Without you, it would have been impossible to perform all the work in time. My friends helped me to enjoy my life outside the university. We had 'borrels', dinners, vacations (peregrino!), played tennis, discovered San Francisco and its beautiful surroundings, communicated via msn, email, the telephone, and skype. And, you always had a listening ear when I needed it, thanks for that! Egon and Marije, thanks for being my paranymphs.

I thank my parents for all the support and motivation over the years. Mum, I appreciate your attempts to understand what exactly your daughter was doing at the VU, and explaining that to others. Dad, I appreciate your critical comments on my work, and our conversations on more than just embodied agents. And last but certainly not least, I thank Erik. You have definitely learned about the good and the bad of the life of a PhD student. Thanks for always being there for me!

Curriculum Vitae

Henriette C. van Vugt werd geboren op 9 juni 1977 te Tilburg. In 1995 voltooide zij de VWO-opleiding aan het Theresialyceum te Tilburg (cum laude), waarna ze gedurende een jaar in Granada de Spaanse taal leerde beheersen.

In 1996 startte zij de studie Taal, Spraak en Informatica aan de Radboud Universiteit Nijmegen, en in 1997 stroomde zij door naar de bovenbouwstudie Cognitiewetenschap. Tijdens haar studie bracht zij vier maanden door aan de Universidade de Coimbra in Portugal. Zij specialiseerde zich in de cognitieve ergonomie en in 2001 studeerde zij af onder supervisie van Dr. Eduard Hoenkamp.

Hierna volgde zij de master Human-Centered Computing Systems aan de University of Sussex in Engeland, die zij in 2002 cum laude afrondde. Haar afstudeerwerk betrof het ontwikkelen en evalueren van een 'awareness' systeem voor thuisgebruik, en werd uitgevoerd aan de TU/Eindhoven onder supervisie van Dr. Panos Markopoulos en Dr. Ian Wakeman.

Van januari 2003 tot maart 2004 werkte zij als junior docent Informatiekunde aan de Universiteit van Utrecht.

Van maart 2004 tot maart 2008 deed zij als promovenda onderzoek aan de Vrije Universiteit in Amsterdam, onder supervisie van Prof. Dr. Jan Kleinnijenhuis (Faculteit der Sociale Wetenschappen), Prof. Dr. Gerrit van der Veer (Faculteit der Exacte Wetenschappen), Dr. mult. Johan Hoorn, and Dr. Elly Konijn. Ruim een jaar was zij lid van de Commissie Wetenschap van de Faculteit der Sociale Wetenschappen. In het kader van haar promotie deed zij gedurende drie maanden onderzoek in het Virtual Human Interaction Lab aan Stanford University in de VS, waar zij samenwerkte met dr. Jeremy Bailenson.

Per 1 maart 2008 is Henriette van Vugt aangesteld als senior onderzoeker bij Philips Research te Eindhoven.

SIKS dissertation series

1998-1	Johan van den Akker (CWI)	DEGAS - An Active, Temporal Database of Autonomous Objects
1998-2	Floris Wiesman (UM)	Information Retrieval by Graphically Browsing Meta-Information
1998-3	Ans Steuten (TUD)	A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective
1998-4	Dennis Breuker (UM)	Memory versus Search in Games
1998-5	E.W.Oskamp (RUL)	Computerondersteuning bij Straftoemeting
1999-1	Mark Sloof (VU)	Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products
1999-2	Rob Potharst (EUR)	Classification using decision trees and neural nets
1999-3	Don Beal (UM)	The Nature of Minimax Search
1999-4	Jacques Penders (UM)	The practical Art of Moving Physical Objects
1999-5	Aldo de Moor (KUB)	Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems
1999-6	Niek J.E. Wijngaards (VU)	Re-design of compositional systems
1999-7	David Spelt (UT)	Verification support for object database design
1999-8	Jacques H.J. Lenting (UM)	Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation
2000-1	Frank Niessink (VU)	Perspectives on Improving Software Maintenance
2000-2	Koen Holtman (TUE)	Prototyping of CMS Storage Management
2000-3	Carolien M.T. Metselaar (UVA)	Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief
2000-4	Geert de Haan (VU)	ETAG, A Formal Model of Competence Knowledge for User Interface Design
2000-5	Ruud van der Pol (UM)	Knowledge-based Query Formulation in Information Retrieval
2000-6	Rogier van Eijk (UU)	Programming Languages for Agent Communication
2000-7	Niels Peek (UU)	Decision-theoretic Planning of Clinical Patient Management
2000-8	Veerle Coup (EUR)	Sensitivity Analyis of Decision-Theoretic Networks

2000-9	Florian Waas (CWI)	Principles of Probabilistic Query Optimization
2000-10	Niels Nes (CWI)	Image Database Management System Design Considerations, Algorithms and Architecture
2000-11	Jonas Karlsson (CWI)	Scalable Distributed Data Structures for Database Management
2001-1	Silja Renooij (UU)	Qualitative Approaches to Quantifying Probabilistic Networks
2001-2	Koen Hindriks (UU)	Agent Programming Languages: Programming with Mental Models
2001-3	Maarten van Someren (UvA)	Learning as problem solving
2001-4	Evgueni Smirnov (UM)	Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets
2001-5	Jacco van Ossenbruggen (VU)	Processing Structured Hypermedia: A Matter of Style
2001-6	Martijn van Welie (VU)	Task-based User Interface Design
2001-7	Bastiaan Schonhage (VU)	Diva: Architectural Perspectives on Information Visualization
2001-8	Pascal van Eck (VU)	A Compositional Semantic Structure for Multi-Agent Systems Dynamics
2001-9	Pieter Jan 't Hoen (RUL)	Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes
2001-10	Maarten Sierhuis (UvA)	Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design
2001-11	Tom M. van Engers (VUA)	Knowledge Management: The Role of Mental Models in Business Systems Design
2002-01	Nico Lassing (VU)	Architecture-Level Modifiability Analysis
2002-02	Roelof van Zwol (UT)	Modelling and searching web-based document collections
2002-03	Henk Ernst Blok (UT)	Database Optimization Aspects for Information Retrieval
2002-04	Juan Roberto Castelo Valdueza (UU)	The Discrete Acyclic Digraph Markov Model in Data Mining
2002-05	Radu Serban (VU)	The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents
2002-06	Laurens Mommers (UL)	Applied legal epistemology; Building a knowledge-based ontology of the legal domain
2002-07	Peter Boncz (CWI)	Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications
2002-08	Jaap Gordijn (VU)	Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas
2002-09	Willem-Jan van den Heuvel (KUB)	Integrating Modern Business Applications with Objectified Legacy Systems
2002-10	Brian Sheppard (UM)	Towards Perfect Play of Scrabble
2002-11	Wouter C.A. Wijngaards (VU)	Agent Based Modelling of Dynamics: Biological and Organisational Applications
2002-12	Albrecht Schmidt (Uva)	Processing XML in Database Systems
2002-13	Hongjing Wu (TUE)	A Reference Architecture for Adaptive Hypermedia Applications
2002-14	Wieke de Vries (UU)	Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems
2002-15	Rik Eshuis (UT)	Semantics and Verification of UML Activity Diagrams for Workflow Modelling
2002-16	Pieter van Langen (VU)	The Anatomy of Design: Foundations, Models and Applications

2002-17	Stefan Manegold (UVA)	Understanding, Modeling, and Improving Main-Memory Database Performance
2003-01	Heiner Stuckenschmidt (VU)	Ontology-Based Information Sharing in Weakly Structured Environments
2003-02	Jan Broersen (VU)	Modal Action Logics for Reasoning About Reactive Systems
2003-03	Martijn Schuemie (TUD)	Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy
2003-04	Milan Petkovic (UT)	Content-Based Video Retrieval Supported by Database Technology
2003-05	Jos Lehmann (UVA)	Causation in Artificial Intelligence and Law - A modelling approach
2003-06	Boris van Schooten (UT)	Development and specification of virtual environments
2003-07	Machiel Jansen (UvA)	Formal Explorations of Knowledge Intensive Tasks
2003-08	Yongping Ran (UM)	Repair Based Scheduling
2003-09	Rens Kortmann (UM)	The resolution of visually guided behaviour
2003-10	Andreas Lincke (UvT)	Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture
2003-11	Simon Keizer (UT)	Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks
2003-12	Roeland Ordelman (UT)	Dutch speech recognition in multimedia information retrieval
2003-13	Jeroen Donkers (UM)	Nosce Hostem - Searching with Opponent Models
2003-14	Stijn Hoppenbrouwers (KUN)	Freezing Language: Conceptualisation Processes across ICT-Supported Organisations
2003-15	Mathijs de Weerdt (TUD)	Plan Merging in Multi-Agent Systems
2003-16	Menzo Windhouwer (CWI)	Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses
2003-17	David Jansen (UT)	Extensions of Statecharts with Probability, Time, and Stochastic Timing
2003-18	Levente Kocsis (UM)	Learning Search Decisions
2004-01	Virginia Dignum (UU)	A Model for Organizational Interaction: Based on Agents, Founded in Logic
2004-02	Lai Xu (UvT)	Monitoring Multi-party Contracts for E-business
2004-03	Perry Groot (VU)	A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
2004-04	Chris van Aart (UVA)	Organizational Principles for Multi-Agent Architectures
2004-05	Viara Popova (EUR)	Knowledge discovery and monotonicity
2004-06	Bart-Jan Hommes (TUD)	The Evaluation of Business Process Modeling Techniques
2004-07	Elise Boltjes (UM)	Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
2004-08	Joop Verbeek (UM)	Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise
2004-09	Martin Caminada (VU)	For the Sake of the Argument; explorations into argument-based reasoning
2004-10	Suzanne Kabel (UVA)	Knowledge-rich indexing of learning-objects
2004-11	Michel Klein (VU)	Change Management for Distributed Ontologies
2004-12	The Duy Bui (UT)	Creating emotions and facial expressions for embodied agents
2004-13	Wojciech Jamroga (UT)	Using Multiple Models of Reality: On Agents who Know how to Play

2004-14	Paul Harrenstein (UU)	Logic in Conflict. Logical Explorations in Strategic Equilibrium
2004-15	Arno Knobbe (UU)	Multi-Relational Data Mining
2004-16	Federico Divina (VU)	Hybrid Genetic Relational Search for Inductive Learning
2004-17	Mark Winands (UM)	Informed Search in Complex Games
2004-18	Vania Bessa Machado (UvA)	Supporting the Construction of Qualitative Knowledge Models
2004-19	Thijs Westerveld (UT)	Using generative probabilistic models for multimedia retrieval
2004-20	Madelon Evers (Nyenrode)	Learning from Design: facilitating multidisciplinary design teams
2005-01	Floor Verdenius (UVA)	Methodological Aspects of Designing Induction-Based Applications
2005-02	Erik van der Werf (UM))	AI techniques for the game of Go
2005-03	Franc Grootjen (RUN)	A Pragmatic Approach to the Conceptualisation of Language
2005-04	Nirvana Meratnia (UT)	Towards Database Support for Moving Object data
2005-05	Gabriel Infante-Lopez (UVA)	Two-Level Probabilistic Grammars for Natural Language Parsing
2005-06	Pieter Spronck (UM)	Adaptive Game AI
2005-07	Flavius Frasincar (TUE)	Hypermedia Presentation Generation for Semantic Web Information Systems
2005-08	Richard Vdovjak (TUE)	A Model-driven Approach for Building Distributed Ontology-based Web Applications
2005-09	Jeen Broekstra (VU)	Storage, Querying and Inferencing for Semantic Web Languages
2005-10	Anders Bouwer (UVA)	Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments
2005-11	Elth Ogston (VU)	Agent Based Matchmaking and Clustering - A Decentralized Approach to Search
2005-12	Csaba Boer (EUR)	Distributed Simulation in Industry
2005-13	Fred Hamburg (UL)	Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen
2005-14	Borys Omelayenko (VU)	Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics
2005-15	Tibor Bosse (VU)	Analysis of the Dynamics of Cognitive Processes
2005-16	Joris Graaumans (UU)	Usability of XML Query Languages
2005-17	Boris Shishkov (TUD)	Software Specification Based on Re-usable Business Components
2005-18	Danielle Sent (UU)	Test-selection strategies for probabilistic networks
2005-19	Michel van Dartel (UM)	Situated Representation
2005-20	Cristina Coteanu (UL)	Cyber Consumer Law, State of the Art and Perspectives
2005-21	Wijnand Derks (UT)	Improving Concurrency and Recovery in Database Systems by Exploiting Application
	Semantics	
2006-01	Samuil Angelov (TUE)	Foundations of B2B Electronic Contracting
2006-02	Cristina Chisalita (VU)	Contextual issues in the design and use of information technology in organizations
2006-03	Noor Christoph (UVA)	The role of metacognitive skills in learning to solve problems
2006-04	Marta Sabou (VU)	Building Web Service Ontologies
2006-05	Cees Pierik (UU)	Validation Techniques for Object-Oriented Proof Outlines
2006-06	Ziv Baida (VU)	Software-aided Service Bundling - Intelligent Methods and Tools for Graphical Service Modeling

2006-07	Marko Smiljanic (UT)	XML schema matching – balancing efficiency and effectiveness by means of clustering
2006-08	Eelco Herder (UT)	Forward, Back and Home Again - Analyzing User Behavior on the Web
2006-09	Mohamed Wahdan (UM)	Automatic Formulation of the Auditor's Opinion
2006-10	Ronny Siebes (VU)	Semantic Routing in Peer-to-Peer Systems
2006-11	Joeri van Ruth (UT)	Flattening Queries over Nested Data Types
2006-12	Bert Bongers (VU)	Interactivation - Towards an e-cology of people, our technological environment, and the arts
2006-13	Henk-Jan Lebbink (UU)	Dialogue and Decision Games for Information Exchanging Agents
2006-14	Johan Hoorn (VU)	Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change
2006-15	Rainer Malik (UU)	CONAN: Text Mining in the Biomedical Domain
2006-16	Carsten Riggelsen (UU)	Approximation Methods for Efficient Learning of Bayesian Networks
2006-17	Stacey Nagata (UU)	User Assistance for Multitasking with Interruptions on a Mobile Device
2006-18	Valentin Zhizhkun (UVA)	Graph transformation for Natural Language Processing
2006-19	Birna van Riemsdijk (UU)	Cognitive Agent Programming: A Semantic Approach
2006-20	Marina Velikova (UvT)	Monotone models for prediction in data mining
2006-21	Bas van Gils (RUN)	Aptness on the Web
2006-22	Paul de Vrieze (RUN)	Fundaments of Adaptive Personalisation
2006-23	Ion Juvina (UU)	Development of Cognitive Model for Navigating on the Web
2006-24	Laura Hollink (VU)	Semantic Annotation for Retrieval of Visual Resources
2006-25	Madalina Drugan (UU)	Conditional log-likelihood MDL and Evolutionary MCMC
2006-26	Vojkan Mihajlovic (UT)	Score Region Algebra: A Flexible Framework for Structured Information Retrieval
2006-27	Stefano Bocconi (CWI)	Vox Populi: generating video documentaries from semantically annotated media repositories
2006-28	Borkur Sigurbjornsson (UVA)	Focused Information Access using XML Element Retrieval
2007-01	Kees Leune (UvT)	Access Control and Service-Oriented Architectures
2007-02	Wouter Teepe (RUG)	Reconciling Information Exchange and Confidentiality: A Formal Approach
2007-03	Peter Mika (VU)	Social Networks and the Semantic Web
2007-04	Jurriaan van Diggelen (UU)	Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach
2007-05	Bart Schermer (UL)	Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance
2007-06	Gilad Mishne (UVA)	Applied Text Analytics for Blogs
2007-07	Natasa Jovanovic' (UT)	To Whom It May Concern - Addressee Identification in Face-to-Face Meetings
2007-08	Mark Hoogendoorn (VU)	Modeling of Change in Multi-Agent Organizations
2007-09	David Mobach (VU)	Agent-Based Mediated Service Negotiation
2007-10	Huib Aldewereld (UU)	Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols
2007-11	Natalia Stash (TUE)	Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System

2007-12	Marcel van Gerven (RUN)	Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty
2007-13	Rutger Rienks (UT)	Meetings in Smart Environments; Implications of Progressing Technology
2007-14	Niek Bergboer (UM)	Context-Based Image Analysis
2007-15	Joyca Lacroix (UM)	NIM: a Situated Computational Memory Model
2007-16	Davide Grossi (UU)	Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems
2007-17	Theodore Charitos (UU)	Reasoning with Dynamic Networks in Practice
2007-18	Bart Orriens (UvT)	On the development an management of adaptive business collaborations
2007-19	David Levy (UM)	Intimate relationships with artificial partners
2007-20	Slinger Jansen (UU)	Customer Configuration Updating in a Software Supply Network
2007-21	Karianne Vermaas (UU)	Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005
2007-22	Zlatko Zlatev (UT)	Goal-oriented design of value and process models from patterns
2007-23	Peter Barna (TUE)	Specification of Application Logic in Web Information Systems
2007-24	Georgina Ramrez Camps (CWI)	Structural Features in XML Retrieval
2007-25	Joost Schalken (VU)	Empirical Investigations in Software Process Improvement
2008-01	Katalin Boer-Sorbn (EUR)	Agent-Based Simulation of Financial Markets: A modular, continuous-time approach
2008-02	Alexei Sharpanskykh (VU)	On Computer-Aided Methods for Modeling and Analysis of Organizations
2008-03	Vera Hollink (UVA)	Optimizing hierarchical menus: a usage-based approach
2008-04	Ander de Keijzer (UT)	Management of Uncertain Data - towards unattended integration
2008-05	Bela Mutschler (UT)	Modeling and simulating causal dependencies on process-aware information systems from a cost perspective
2008-06	Arjen Hommersom (RUN)	On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective
2008-07	Peter van Rosmalen (OU)	Supporting the tutor in the design and support of adaptive e-learning
2008-08	Janneke Bolt (UU)	Bayesian Networks: Aspects of Approximate Inference
2008-09	Christof van Nimwegen (UU)	The paradox of the guided user: assistance can be counter-effective
2008-10	Wauter Bosma (UT)	Discourse oriented summarization
2008-11	Vera Kartseva (VU)	Designing Controls for Network Organizations: A Value-Based Approach
2008-12	Jozsef Farkas (RUN)	A Semiotically Oriented Cognitive Model of Knowledge Representation
2008-13	Caterina Carraciolo (UVA)	Topic Driven Access to Scientific Handbooks
2008-14	Arthur van Bunningen (UT)	Context-Aware Querying; Better Answers with Less Effort
2008-15	Martijn van Otterloo (UT)	The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains

Dissertation series, faculty of Social Sciences, VU University Amsterdam

1 M.M. Meijer, (2004), Does Success Breed Success? Effects of news and advertising on corporate reputation. Askant Academic Publishers: Amsterdam, isbn 90 5260 145 3.

2 J.J. Woldendorp, (2005), The Polder Model: From Disease to Miracle? Dutch Neo-corporatism 1965-2000. Amsterdam, isbn 90 77472 05 3.

3 Y.P. Ongena, (2005), Interviewer and Respondent. Interaction in Survey Interviews, Yfke Ongena. Amsterdam, isbn 90 9020070 3.

4 M.E.D. Lamboo, (2005), Integriteitsbeleid van de Nederlandse politie. Eburon: Delft, isbn 90 5972 088 L.

5 T.E. Aalberts, (2006), Politics of Sovereignty. Vrije Universiteit Amsterdam, isbn 10: 90 9020656 6 / isbn 13: 978 90 9020656 1.

6 E. Zwart, (2006), In Pursuit of Comfort. The Transnationalisation Process of Malaysian Chinese Small and Medium Enterprises. Vrije Universiteit Amsterdam, isbn 90 8659 024 1.

7 H.Haenen, (2006), Afrikaans Denken. Ontmoeting, dialoog en frictie. Een filosofisch onderzoek. Buijten en Schipperheijn Motief, Amsterdam.

8 A. Claver, (2006), Commerce and capital in colonial Java. Trade finance and commercial relations between Europeans and Chinese, 1820s-1942. Vrije Universiteit Amsterdam.

9 S. van der Pas, (2006), Intergenerational relationships of older adults. Family structure, contact and norms. Vrije Universiteit Amsterdam, isbn 10: 90 9022075 X / isbn13: 978 90 9021075 9.

10 P. Mika, (2007), Social Netwoks and the Semantic Web. SIKS dissertation series No. 2007-03.

11 K.E. Knibbe, (2007), Faith in the Familiar. Continuity and Change in Religious Practices and Moral Orientations of the South of Limburg, The Netherlands. Vrije Universiteit Amsterdam.

12 C.L. Carabain, (2007), Taking Too Much for Granted? A Study on the Measure of Social Attitudes. Gildeprint Drukkerijen B.V., Enschede, isbn 978 90 8659 075 9.

13 L.M. Cruz, (2007), Repression and Rebellions in Southern Mexico. The search for a political economy of dignity. Dutch University Press, Amsterdam, isbn 978 90 361 0050 2.

14 C.E. de Vries, (2007), European Integration and National Elections. The impact of EU Issue Voting on National Electoral Politics. Vrije Universiteit Amsterdam, isbn 13: 978 90 9022112 0.

15 E. Kolthoff, (2007), Ethics and New Public Management. Empirical Research into Effects of Businesslike Government on Ethics and Integrity. Boom Juridische Uitgevers, isbn 978 90 5454 905 5.

16 R. Vliegenthart, (2007), Framing Immigration and Integration. Facts, Parliament, Media and Anti-Immigrant Party Support in the Netherlands. Vrije Universiteit Amsterdam, isbn 978 90 9022279 0.

17 L.L.J. van Noije, (2007), The Democratic Deficit Closer to Home. Agendabuilding relations between parliament and the press, and the impact of European integration, in the United Kingdom, the Netherlands and France. Vrije Universiteit Amsterdam, isbn 978-90-8659-161-9.

18 R. de Man, (2007), Where to draw the line? Antecedents and consequences of the integration of work in home life. Vrije Universiteit Amsterdam.

19 F. Krijtenburg, (2007), Cultural Ideologies of Peace and Conflict. Vrije Universiteit, isbn 978 90 8659 167 1.

20 R. van Steden, (2007), Privatizing Policing. Describing and explaining the growth of private security. BJU Legal Publishers, isbn 978 90 5454 953 6.

21 L. Reijn, (2008), Identity at work. Control and commitment in postbureaucratic organisations.Vrije Universiteit Amsterdam, isbn 978 90 8659 180 0.

22 B. Vis, (2008), Biting the Bullet or Steering Clear? Politics of (Not-) Unpopular Welfare State Reform in Advanced Capitalist Democracies. Vrije Universiteit Amsterdam, isbn 978 90 9022715 3.

23 G. van Noort, (2008), Bricks versus Clicks. A self-regulation perspective on consumer responses in online and conventional shopping environment. Vrije Universiteit Amsterdam, 978-90-9022845-7.

24 M.J.M. de Koning (2008), Zoeken naar een ‘zuivere’ islam. Geloofsbeving en identiteisvorming van jonge Marokkaans-Nederlandse moslims. Bert Bakker.

25 P. Wiepking (2008), For the Love of Mankind. A Sociological Study on Charitable Giving. Labor Grafimedia, isbn 978 90 77383 07 0.

26 W. Stam, (2008), Does Networking Work? A Study of Social Capital and Performance Among New Ventures in the Dutch Open Source Software Industry. Vrije Universiteit Amsterdam, isbn 978-90-9023014-6.